

Klasterisasi Status Keberlanjutan Karyawan Kontrak Menggunakan K-Medoids

Clustering Contract Employees' Sustainability Status Using K-Medoids

Dinda Ramadhaniaty¹, Novianti Puspitasari^{2*}, Anindita Septiarini³

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman

^{1,2,3}Jalan Sambaliung, No. 9, Kampus Gunung Kelua, Samarinda 75119, Indonesia

dindaramadhaniaty975@gmail.com¹, novia.ftik.unmul@gmail.com^{2*}, anindita.septiarini@gmail.com³

Abstrak – Karyawan kontrak adalah karyawan tidak tetap yang dipekerjakan untuk melakukan pekerjaan di dalam perusahaan, dan tidak ada jaminan kelangsungan masa kerjanya. Penilaian terhadap karyawan tetap maupun karyawan kontrak dilakukan dengan cara menilai kinerja seorang karyawan tersebut. Oleh karena itu, perusahaan selalu melakukan evaluasi terhadap kinerja karyawannya. Penelitian ini bertujuan untuk mengelompokkan status karyawan kontrak yang layak diperpanjang atau tidak layak secara objektif sehingga mempermudah perusahaan dalam menganalisis data agar hasil lebih tepat dan dapat menghindari kecurangan dalam prosesnya. Algoritma yang digunakan dalam penelitian ini adalah K-Medoids menggunakan tiga metode pengukuran jarak yaitu Euclidean Distance, Manhattan Distance dan Chebyshev Distance. Sementara, untuk uji akurasi menggunakan Silhouette Coefficient (SC) dan Sum of Square Error (SSE). Data yang digunakan adalah data penilaian karyawan kontrak berjumlah 42 karyawan kontrak. Berdasarkan pengujian nilai SC diperoleh bahwa dari ketiga metode pengukuran jarak yang digunakan, chebyshev distance menghasilkan nilai yang mendekati 1 dengan nilai 0,3220214. Sedangkan, hasil uji cluster SSE diperoleh bahwa nilai error terkecil dimiliki oleh 2 cluster. Lebih lanjut, hasil pengelompokkan status keberlanjutan karyawan kontrak menempatkan 20 data ke dalam kategori Tidak Layak (C1) dan 22 data ke dalam kategori Layak (C2).

Kata Kunci: Karyawan, K-Medoids, Chebyshev Distance, SSE, Silhouette Coefficient.

Abstract – Contract employees are temporary workers hired to perform tasks within a company without any guarantee of tenure. Evaluations of both permanent and contract employees are conducted based on their performance, which is essential for the company's continuous performance assessment. This study aims to classify the status of contract employees to determine their eligibility for contract extension objectively. This objective classification facilitates the company's data analysis, resulting in more accurate outcomes and minimizing the risk of fraud in the evaluation process. The algorithm employed in this research is K-medoids, utilizing three distance measurement methods: Euclidean distance, Manhattan distance, and Chebyshev distance. The accuracy of the classifications was tested using the silhouette coefficient (SC) and sum of squares error (SSE). The data set comprised performance assessments of 42 contract employees. Testing the SC values revealed that among the three distance measurement methods, the Chebyshev distance yielded the closest value to 1, specifically 0.3220214. Additionally, the SSE cluster test results indicated that the smallest error value was achieved with 2 clusters. Consequently, the grouping results classified 20 employees as Ineligible (C1) and 22 employees as Eligible (C2) for contract extension.

Keywords: Employee, K-Medoids, Chebyshev Distance, SSE, Silhouette Coefficient.

1. Pendahuluan

Karyawan merupakan sumber daya manusia (SDM) yaitu aset yang sangat berharga dalam sebuah perusahaan maupun instansi dan harus dikelola dengan baik agar dapat memberikan kontribusi yang terbaik bagi kemajuan perusahaan [1]. Pada sebuah perusahaan biasanya terdapat karyawan tetap dan karyawan kontrak [2]. Karyawan tetap merupakan karyawan yang telah memiliki kontrak ataupun perjanjian kerja dalam jangka waktu yang tidak ditetapkan (permanen) selama karyawan bekerja dengan baik [3]. Karyawan kontrak adalah karyawan tidak tetap yang dipekerjakan untuk melakukan pekerjaan di dalam perusahaan dan tidak ada jaminan kelangsungan masa kerjanya [4]. Keberlanjutan kerja dari karyawan kontrak sangat ditentukan oleh kinerjanya, jika kinerjanya baik maka karyawan kontrak akan tetap tinggal di perusahaan dengan penambahan masa kontrak, namun jika kinerjanya tidak baik maka perusahaan tidak akan melanjutkan kontrak karyawan tersebut [5]. Penilaian terhadap karyawan tetap maupun karyawan kontrak dilakukan dengan cara menilai kinerja seorang karyawan tersebut. Oleh karena itu, setiap perusahaan atau pimpinan lembaga selalu melakukan evaluasi terhadap kinerja karyawannya. Setiap tahun ada karyawan yang diperpanjang ataupun tidak diperpanjang kontrak kerjanya. Dalam hal menentukan status seorang karyawan diperpanjang atau tidak diperpanjang kontrak kerjanya sering terjadi kesulitan dalam menentukan penilaian serta memerlukan waktu dan proses yang lama. Kesulitan tersebut meliputi banyaknya kriteria penilaian yang diperlukan dan penilaian yang dilakukan tidak profesional. Hal tersebut dapat menimbulkan ketidakadilan bagi karyawan sehingga dapat memicu dampak negatif bagi perusahaan serta dapat menurunkan kualitas perusahaan [6]. Berdasarkan permasalahan yang ada perlu adanya pendekatan yang lebih objektif dalam melakukan proses penentuan status perpanjangan karyawan kontrak agar dapat meningkatkan kualitas sebuah perusahaan tanpa dipengaruhi oleh subjektivitas penilaian pimpinan.

Penentuan status karyawan kontrak diperpanjang atau tidak, dapat menggunakan salah satu metode pada data mining yaitu *clustering*. *Clustering* adalah proses membagi sekumpulan objek data menjadi himpunan bagian yang disebut *cluster*. Objek-objek dalam suatu *cluster* memiliki sifat yang mirip satu sama lain dan berbeda dengan *cluster* lainnya [7],[8]. Oleh karena itu, pengelompokan sangat berguna dan dapat menemukan grup atau grup yang tidak dikenal dalam data [9]. *Clustering* memiliki berbagai macam algoritma, salah satu algoritma *clustering* yang dapat digunakan adalah algoritma K-Medoids [10]. K-Medoids adalah algoritma yang dapat digunakan untuk mendapatkan medoids didalam sebuah kelompok (*cluster*) yang merupakan titik pusat dari suatu kelompok [11]. Algoritma ini telah digunakan untuk pengelompokan data gaji karyawan [12]. Algoritma K-Medoids juga digunakan dalam pengelompokan kasus penyakit *acquired immunodeficiency syndrome* (AIDS) berdasarkan provinsi [13], K-Medoid digunakan untuk *clustering* penggunaan jaringan internet [14] dan menentukan penerima program bidikmisi [15]. 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 [16]. Penelitian ini bertujuan agar dapat membantu dalam proses menentukan perpanjangan status kontrak karyawan dan mengklasterisasikan status karyawan kontrak yang layak diperpanjang kontraknya atau tidak diperpanjang

2. Metode Penelitian

2.1. K-Medoids

K-Medoids atau *Partitioning Around Medoids* (PAM), perwakilan (medoid) adalah algoritma *clustering* yang menentukan pusat *cluster* untuk setiap *cluster* menggunakan sebuah objek. Medoid adalah objek *cluster* yang paling sentral dan memiliki jarak paling dekat ke titik lain. Metode K-Medoids dalam menentukan jumlah k awal secara random digunakan metode elbow [17]. Metode K-Medoids sangat efisien untuk kumpulan data kecil. Di dalam perhitungan algoritma K-Medoids, hal paling penting adalah mencari titik yang paling representatif (medoids) dalam sebuah kumpulan data dengan menghitung jarak dalam kelompok dari semua kemungkinan

kombinasi dari medoids sehingga jarak antar titik dalam suatu *cluster* kecil sedangkan jarak titik antar *cluster* besar [18].

2.2. Euclidean Distance

Euclidean distance adalah salah satu metode perhitungan jarak yang paling sering digunakan yaitu perhitungan untuk mengukur jarak dua titik dalam *euclidean space* yang mempelajari hubungan antara sudut dan jarak. Rumus persamaan untuk menghitung euclidean distance dalam persamaan (1) [19], [20].

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1)$$

dengan d adalah jarak antara data dengan medoid, i adalah data ke- i , n adalah jumlah data, x_i adalah data pada setiap data ke- i dan y_i adalah data pada pusat *cluster* ke- i .

2.3. Manhattan Distance

Manhattan Distance adalah metode perhitungan jarak untuk menghitung perbedaan nilai absolut (mutlak) antara koordinat sepasang objek [21]. Rumus persamaan yang dapat digunakan untuk menghitung jarak dengan *Manhattan Distance* dapat dilihat pada persamaan (2) [22].

$$d_{\text{Manhattan}}(x, y) = \sum_i^n |x_i - y_i| \quad (2)$$

dengan d adalah jarak antara data dengan medoid, i adalah data ke- i , n adalah jumlah data x_i adalah data pada setiap data ke- i dan y_i adalah data pada pusat klaster ke- i .

2.4. Chebyshev Distance

Chebyshev Distance adalah sebuah metode pengukuran jarak berdasarkan nilai absolut atau nilai mutlak perbedaan mutlak antara dua titik koordinat. Jika terdapat dua vektor masing-masing dengan nilai yang berbeda elemen, jarak yang diukur oleh *Chebyshev* berdasarkan nilai absolut dari perbedaan antara elemen vektor dan jumlah data harus sama secara otomatis [23]. Rumus persamaan yang digunakan untuk menghitung jarak dengan *chebyshev distance* terdapat di persamaan (3) [24].

$$d_{\text{Chebyshev}}(x, y) = \max_{i=1}^n |x_i - y_i| \quad (3)$$

dengan d adalah jarak antara data dengan medoid, i adalah data ke- i , n adalah jumlah data x_i adalah data pada setiap data ke- i dan y_i adalah data pada pusat klaster ke- i .

2.5. Silhouette Coefficient

Silhouette Coefficient merupakan salah satu metode validasi yang digunakan untuk melihat kualitas dan kekuatan *cluster* atau hasil kelompok setiap perhitungan jarak, yaitu seberapa baik objek ditempatkan di *cluster* [25]. Rumus menghitung nilai *silhouette coefficient* dapat dilihat pada persamaan (4) [19].

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

dengan, $a_{(i)}$ adalah rata-rata jarak suatu objek ke- i dengan semua objek lain pada satu *cluster*; $b_{(i)}$ adalah rata-rata jarak dari objek ke- i dengan semua objek pada *cluster* lain; $S_{(i)}$ adalah nilai *Silhouette Coefficient*.

Nilai *Silhouette* yang dihitung dapat memberikan hasil yang bervariasi antara -1 hingga 1. $S_i = 1$ artinya data tersebut berada di dalam *cluster* yang benar, jika $S_i = 0$ maka data berada di antara kedua *cluster*, sehingga data tersebut tidak jelas harus dimasukkan ke dalam *cluster* yang mana dan jika $S_i = -1$ data mengalami *overlapping* sehingga data tersebut lebih tepat dimasukan ke dalam *cluster* yang terpisah [26].

2.6. Sum of Square Error (SEE)

Sum of Square Error (SSE) merupakan cara dalam validasi *cluster* melalui jumlah kuadrat setiap anggota *cluster* menuju pusatnya. Semakin jauh jarak yang membentuk titik siku, maka hasil *cluster* tersebut menjadi yang paling optimal. Prinsip perhitungan nilai SSE yaitu *trial and error* dan akan terus diulang hingga nilai *cluster* mencapai nilai minimum dengan menggunakan persamaan (5) [27], [28].

$$SSE = \sum_{k=1}^k \sum_{i=1}^n (p_i - m_k)^2 \quad (5)$$

SSE adalah *Sum of Square Error*, i adalah data ke- i , k adalah jumlah klaster, n yaitu jumlah data, p_i yaitu data pada setiap data ke- i dan m_i adalah data pada pusat klaster ke- i .

2.7. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data primer yang diperoleh dengan melakukan penelitian secara langsung di Bankaltimtara Kantor Cabang Tanjung Redeb dan juga data sekunder dari hasil studi dokumentasi pada *website* penilaian karyawan Bankaltimtara. Dalam hal ini data yang diperoleh berjumlah 42 orang, berasal dari hasil penilaian karyawan kontrak tahun 2022. Data dari setiap *record* terdiri dari hasil penilaian dari pihak yang melakukan penilaian seperti terlihat pada Tabel 1.

Tabel 1. Data penilaian karyawan kontrak 2022.

Pegawai 1						
No	Faktor penilaian	Hasil penilaian		Hasil koefisien		Hasil
		I	II	I-A	II-A	
1	Potensi	100	100	40	60	100
2	Kinerja	120	100	48	60	108
3	Kerjasama	70	50	28	30	58
4	Kemandirian	60	50	24	30	54
5	Pengalaman	60	40	24	24	48
6	Perilaku	70	50	28	30	58
7	Ketelitian	25	25	10	15	25
8	Fleksibilitas	30	25	12	15	27
9	Stabilitas Emosi	40	25	16	15	31
10	Keuletan	30	25	12	15	27
11	Jumlah Penilaian	605	490	242	294	536
12	Koefisien	40%	60%			
.
.
.
.
Pegawai 42						
No	Faktor penilaian	Hasil penilaian		Hasil koefisien		Hasil
		I	II	I-A	II-A	
1	Potensi	120	100	48	60	108
2	Kinerja	100	80	40	48	88
3	Kerjasama	60	50	24	30	54
4	Kemandirian	60	50	24	30	54
5	Pengalaman	60	50	24	30	54
6	Perilaku	60	50	24	30	54
7	Ketelitian	35	25	14	15	29
8	Fleksibilitas	30	25	12	15	27
9	Stabilitas Emosi	25	25	10	15	25
10	Keuletan	30	25	12	15	27
11	Jumlah Penilaian	580	480	232	288	520
12	Koefisien	40%	60%			

Berdasarkan Tabel 1 data setiap record terdiri dari hasil penilaian yang berwenang tentang potensi, kinerja, kerjasama, kemandirian, pengalaman, perilaku, ketelitian, fleksibilitas, stabilitas emosi, dan keuletan setiap karyawan kontrak.

3. Hasil dan Pembahasan

3.1. Pengolahan Data

Penelitian ini menggunakan data penilaian karyawan kontrak di Bankaltimtara Kantor Cabang Tanjung Redeb yang berjumlah 42 karyawan kontrak berdasarkan data pada Tabel 1. Selanjutnya tahapan proses *cleaning* data tidak dilakukan karena data yang diperoleh sudah bersih sehingga hanya dilakukan proses *data selection* untuk memilih data dan atribut yang diperlukan. Hasil dari *selection data* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *selection data*.

Nama	Potensi	Kinerja	Kerja sama	Kemandirian	Pengalaman	Perilaku	Ketelitian	Fleksibilitas	Stabilitas emosi	Keuletan
Pegawai 1	100	108	58	54	48	58	25	27	31	27
Pegawai 2	116	116	54	60	44	68	29	27	29	29
Pegawai 3	136	128	60	58	54	60	30	30	30	34
Pegawai 4	108	108	60	54	54	58	23	27	30	27
Pegawai 5	116	116	62	64	64	62	34	29	31	29
Pegawai 6	116	116	62	58	58	60	27	27	29	29
Pegawai 7	100	100	50	60	50	50	30	27	23	28
Pegawai 8	108	96	48	54	54	58	27	31	29	29
Pegawai 9	88	88	50	44	50	50	16	16	16	19
Pegawai 10	96	96	48	54	54	48	24	27	27	24
Pegawai 11	124	108	54	54	48	54	25	27	27	27
Pegawai 12	108	116	54	54	54	58	27	27	27	32
Pegawai 13	72	72	50	46	40	46	20	27	34	25
Pegawai 14	124	108	54	54	58	58	25	27	27	27
Pegawai 15	108	108	54	50	44	54	25	29	30	27
Pegawai 16	116	96	48	54	54	54	29	27	25	27
Pegawai 17	116	116	58	58	48	58	27	27	27	27
Pegawai 18	124	108	64	54	60	58	27	27	32	29
Pegawai 19	108	108	58	48	48	52	27	24	26	24
Pegawai 20	108	108	58	54	48	58	27	29	27	27
Pegawai 21	108	108	54	48	48	58	27	27	29	27
Pegawai 22	116	124	62	58	56	62	27	29	29	26
Pegawai 23	140	140	70	70	60	70	30	35	35	30
Pegawai 24	116	108	54	58	48	54	27	27	27	27
Pegawai 25	120	140	70	60	60	60	30	35	30	35
Pegawai 26	40	40	30	20	20	30	10	15	10	10
Pegawai 27	100	100	60	50	40	60	29	27	29	29
Pegawai 28	80	100	60	50	50	54	23	27	30	27
Pegawai 29	80	88	58	58	54	60	27	27	29	29
Pegawai 30	100	88	48	54	54	58	27	31	29	29
Pegawai 31	88	88	48	54	54	48	24	27	27	24
Pegawai 32	88	88	54	54	54	58	27	27	27	32
Pegawai 33	88	100	54	54	44	58	25	27	27	27
Pegawai 34	108	88	48	54	54	54	29	27	25	27
Pegawai 35	108	108	64	54	60	58	27	27	32	29
Pegawai 36	100	108	58	54	48	58	27	29	27	27
Pegawai 37	108	116	62	58	56	62	27	29	29	26
Pegawai 38	80	80	54	58	48	54	27	27	27	27
Pegawai 39	88	74	64	64	54	64	10	15	10	10
Pegawai 40	88	108	64	54	54	54	30	35	30	35
Pegawai 41	108	108	64	64	64	64	25	27	27	27
Pegawai 42	108	88	54	54	54	54	29	27	25	27

Berdasarkan Tabel 2 hasil *selection data* terdapat 42 data karyawan kontrak dengan 10 atribut yang digunakan dalam penelitian ini yaitu potensi, kinerja, kerjasama, kemandirian, pengalaman, perilaku, ketelitian, fleksibilitas, stabilitas emosi dan keuletan. Data yang digunakan merupakan hasil penilaian dari dua penilai terkait yaitu kepala cabang dan kepala bagian berdasarkan bobot yang telah ditetapkan oleh perusahaan. Jumlah koefisien kedua penilai terkait akan digunakan untuk perhitungan menggunakan algoritma *K-Medoids*.

3.2. Implementasi K-Medoids

Penerapan proses algoritma K-Medoid pada data yang telah melewati proses seleksi data dimulai dari tahapan sebagai berikut.

1) Penentuan Nilai K

Dalam penelitian ini jumlah *cluster* awal atau *k* ditetapkan sebanyak 2 kategori *cluster*.

2) Penentuan Pusat Medoid Awal

Penentuan medoid awal pada proses perhitungan algoritma K-Medoid dilakukan secara acak atau *random* dari data penilaian karyawan kontrak Bankaltimtara Kantor Cabang Tanjung Redeb tahun 2022. Pusat medoid awal yang diambil ditampilkan pada Tabel 3.

Tabel 3. Pusat medoid awal.

Cluster	Potensi	Kinerja	Kerjasama	Kemandirian	...	Fleksibilitas	Stabilitas emosi	Keuletan
C1	108	108	64	64	...	27	27	27
C2	108	88	54	54	...	27	25	27

Berdasarkan Tabel 3 penentuan pusat medoid awal untuk $K=2$ dilakukan dengan mengambil data ke-41 sebagai pusat C1 dan data ke-42 sebagai pusat C2. Pada C1 terdapat atribut Potensi dengan nilai 108, kinerja 108, kerjasama 64, kemandirian 64, pengalaman 64, perilaku 64, ketelitian 25, fleksibilitas 27, stabilitas emosi 27, dan keuletan 27 sedangkan pada C2 atribut Potensi dengan nilai 108, kinerja 88, kerjasama 54, kemandirian 54, pengalaman 54, perilaku 54, ketelitian 29, fleksibilitas 25, stabilitas emosi 25, dan keuletan 27.

3) Penentuan Jarak Terdekat ke Pusat *Cluster*

Setelah menentukan jumlah *cluster* dan pusat medoid awal, proses ini dilakukan dengan cara menghitung jarak terdekat menggunakan data pada Tabel 2. Pada penelitian ini digunakan tiga metode jarak yaitu *Euclidean*, *Manhattan* dan *Chebyshev* menggunakan persamaan (1), persamaan (2) dan persamaan (3) untuk menentukan jarak terpendek ke pusat *cluster* agar dapat melihat perbedaan hasil *clustering*. Pada penelitian ini terdapat 11 variabel sehingga perhitungan dilakukan sebanyak 11 kali terhadap masing-masing variabel. Perhitungan ketiga metode jarak pada iterasi ke 1 ditampilkan pada Tabel 4.

Tabel 4. Hasil perhitungan iterasi ke-1.

Nama	<i>Euclidean</i>		<i>Manhattan</i>		<i>Chebyshev</i>	
	C1	C2	C1	C2	C1	C2
Pegawai 1	22.5	24.2	50	52	16	20
Pegawai 2	26.2	34.7	62	72	20	28
Pegawai 3	38	50.6	90	100	28	40
Pegawai 4	16.3	22.7	35	41	10	20
Pegawai 5	15.5	35.3	37	87	9	28
Pegawai 6	15.2	31.7	40	66	8	28
Pegawai 7	27.8	17.3	72	42	14	12
Pegawai 8	25.8	12.5	64	30	16	8
Pegawai 9	46.6	31.4	141	83	20	20
Pegawai 10	31.8	17.8	80	42	16	12
:	:	:	:	:	:	:
Pegawai 33	33.2	26.1	74	52	20	20
Pegawai 34	31.2	6	72	6	20	6
Pegawai 35	13.6	24.7	29	51	10	20
Pegawai 36	22.4	23.3	50	48	16	20
Pegawai 37	13.6	30.9	33	59	8	28
Pegawai 38	45.4	30.1	100	50	28	28
Pegawai 39	51	43.8	125	127	34	20
Pegawai 40	29.4	32.4	74	72	20	20
Pegawai 41	0	28.6	0	66	0	20
Pegawai 42	28.6	6	66	0	20	0

4) Menentukan Pusat Medoid baru

Dari hasil perhitungan iterasi ke-1, terlihat bahwa nilai kedekatan belum mencapai nilai ideal maka proses perhitungan dilanjutkan ke iterasi berikutnya yaitu menentukan pusat medoid baru dengan cara pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru. Pusat medoid baru yang diambil ditampilkan pada Tabel 5.

Tabel 5. Pusat medoid baru.

Cluster	Potensi	Kinerja	Kerja sama	Kemandirian	...	Fleksibilitas	Stabilitas emosi	Keuletan
C1	108	108	64	64	...	27	27	27
C2	108	88	48	54	...	27	25	27

Berdasarkan Tabel 5 penentuan pusat medoid baru untuk $K=2$ iterasi 2 dilakukan dengan mengambil data ke-41 sebagai pusat C1 dan data ke-34 sebagai pusat C2. Pada C1 terdapat atribut Potensi dengan nilai 108, kinerja 108, kerjasama 64, kemandirian 64, pengalaman 64, perilaku 64, ketelitian 25, fleksibilitas 27, stabilitas emosi 27, dan keuletan 27 sedangkan pada C2 atribut Potensi dengan nilai 108, kinerja 88, kerjasama 48, kemandirian 54, pengalaman 54, perilaku 54, ketelitian 29, fleksibilitas 25, stabilitas emosi 25, dan keuletan 27.

5) Perhitungan nilai *Cost*

Tahapan perhitungan selanjutnya adalah menghitung jarak menggunakan pusat medoid baru, di mana dalam proses ini menghasilkan hasil perhitungan untuk 42 data dengan medoids baru sebagai berikut.

a. *Euclidean Distance*

Hasil pengelompokan menggunakan metode *euclidean distance* iterasi 2 terdapat 19 data pada C1 dan 23 data pada C2 dengan jumlah kedekatan pada iterasi 2 adalah 1042.41073. Langkah selanjutnya adalah menghitung nilai total simpangan (S) dengan menghitung selisih antara *distance* baru – *distance* lama. Jika nilai simpangan negatif < 0 maka ulangi langkah-langkahnya, jika sebaliknya maka terdapat hasil *clustering* dan perhitungan iterasi diberhentikan. Berikut adalah proses perhitungan nilai total simpangan.

$$\begin{aligned} S &= \text{Total cost baru} - \text{Total cost lama} \\ &= 1042,410733 - 1036,800825 \\ &= 5,609908 \end{aligned}$$

Hasil perhitungan pada iterasi ke-1 dan ke-2 diperoleh nilai simpangan yaitu sebesar 5,609908 hasil ini menunjukkan nilai lebih dari 0 maka proses perhitungan berhenti pada iterasi ke-2.

b. *Manhattan Distance*

Hasil pengelompokan menggunakan metode *manhattan distance* iterasi 2 terdapat 19 data pada C1 dan 23 data pada C2 dengan jumlah kedekatan pada iterasi 2 adalah 2356. Langkah selanjutnya adalah menghitung nilai total simpangan (S) dengan menghitung nilai total *distance* baru – *distance* lama. Jika nilai simpangan negatif < 0 maka ulangi langkah-langkahnya, jika sebaliknya maka terdapat hasil *clustering* dan perhitungan iterasi diberhentikan. Berikut adalah proses perhitungan nilai total simpangan.

$$\begin{aligned} S &= \text{Total cost baru} - \text{Total cost lama} \\ &= 2356 - 2320 \\ &= 36 \end{aligned}$$

Dari hasil perhitungan pada iterasi ke-1 dan ke-2 yang telah diperoleh nilai simpangan sebesar 36, hasil ini menunjukkan nilai lebih dari 0 namun terjadi perubahan medoid maka proses perhitungan dilanjutkan hingga tidak terjadi perubahan medoid. Berdasarkan perhitungan jarak menggunakan *Manhattan distance*, proses perhitungan berhenti pada iterasi ke-4.

c. *Chebyshev Distance*

Hasil pengelompokan menggunakan metode *chebyshev distance* iterasi 2 terdapat 33 data pada C1 dan 9 data pada C2 dengan jumlah kedekatan pada iterasi 2 adalah 739. Langkah selanjutnya adalah menghitung nilai total simpangan (S) dengan menghitung nilai selisih *distance* baru – *distance* lama. Jika total *distance* baru – total *distance* lama < 0 maka ulangi langkah-langkahnya, jika sebaliknya maka terdapat hasil *clustering* dan perhitungan iterasi diberhentikan. Berikut adalah proses perhitungan nilai total simpangan.

$$\begin{aligned} S &= \text{Total cost baru} - \text{Total cost lama} \\ &= 739 - 739 \\ &= 0 \end{aligned}$$

Hasil perhitungan pada iterasi ke-1 dan ke-2 yang telah dilakukan didapatkan nilai simpangan yaitu sebesar 0. Hasil ini menunjukkan bahwa proses perhitungan dilanjutkan hingga nilai simpangan lebih dari 0 dan tidak terjadi perubahan medoid. Pada perhitungan jarak menggunakan *chebyshev distance* proses perhitungan berhenti pada iterasi ke-4. Berdasarkan hasil perhitungan jarak medoid menggunakan tiga metode perhitungan jarak yang telah dilakukan sebelumnya, maka diperoleh hasil perhitungan jarak pada iterasi terakhir terlihat pada Tabel 6.

Tabel 6. Hasil perhitungan iterasi terakhir.

Nama	<i>Euclidean</i>		<i>Manhattan</i>		<i>Chebyshev</i>	
	C1	C2	C1	C2	C1	C2
Pegawai 1	22.5	25.8	52	144	20	16
Pegawai 2	26.2	35.2	72	108	28	20
Pegawai 3	38	51.6	100	68	40	28
Pegawai 4	16.3	24.9	41	131	20	10
Pegawai 5	15.5	37.1	87	86	28	9
Pegawai 6	15.2	33.7	66	60	28	8
Pegawai 7	27.8	16.9	42	162	12	14
:	:	:	:	:	:	:
Pegawai 35	13.6	27.7	51	113	20	10
Pegawai 36	22.4	25.1	48	144	20	16
Pegawai 37	13.6	33	59	107	28	8
Pegawai 38	45.4	30.7	50	198	28	28
Pegawai 39	51	45.5	127	227	20	34
Pegawai 40	29.4	34.8	72	138	20	20
Pegawai 41	0	31.2	66	110	20	0
Pegawai 42	28.6	6	0	160	0	20

6) Perhitungan *Silhouette Coefficient*

Metode *Silhouette Coefficient* (SC) digunakan untuk menguji tiga jenis metode pengukuran jarak yaitu *euclidean distance*, *manhattan distance*, dan *chebyshev distance* yang optimal. Hasil perhitungan nilai SC dengan 3 metode pengukuran jarak ditampilkan pada Tabel 7.

Tabel 7. Hasil pengujian *silhouette coefficient*.

<i>Distance</i>	<i>Euclidean</i>	<i>Manhattan</i>	<i>Chebyshev*</i>
<i>Silhouette Coefficient (SC)</i>	0.23328401	0.2642092	0.3220214

Berdasarkan Tabel 7 pengujian nilai SC diperoleh metode pengukuran jarak yang menghasilkan nilai yang paling tinggi atau mendekati 1 yaitu *Chebyshev Distance*.

3.3. Pengujian Cluster

Metode *sum of square error (SSE)* digunakan untuk menguji jumlah *cluster* yang optimal dari perhitungan jarak *Chebyshev Distance*, dimana pengelompokan 2 *cluster* sebelumnya akan dibandingkan dengan *cluster* uji yaitu pengelompokan dengan 3 *cluster* dan 4 *cluster*. Hasil perhitungan nilai SSE dari metode *chebyshev distance* diperoleh hasil perbandingan nilai SSE untuk 2 *cluster*, 3 *cluster*, dan 4 *cluster* yang ditampilkan pada Tabel 8.

Tabel 8. Hasil pengujian SSE.

Cluster	Hasil SSE
K=2	11221.3571
K=3	24659.8333
K=4	25509.7619

Berdasarkan Tabel 8 hasil perhitungan SSE yang disajikan, diperoleh bahwa nilai error terkecil dimiliki oleh 2 *cluster* sedangkan nilai error terbesar dimiliki oleh 4 *cluster*. Hal ini membuktikan bahwa 2 *cluster* adalah *cluster* yang paling optimal yang dapat digunakan dalam penelitian ini karena memiliki nilai *error* terkecil.

3.4. Hasil Clustering

Berdasarkan hasil pengujian cluster menggunakan SSE dan *Silhouette Coefficient* maka jumlah cluster yang sesuai untuk menentukan status keberlanjutan menggunakan K-Medoids adalah 2 cluster dengan metode jarak *Chebyshev Distance*. Hasil penentuan status keberlanjutan karyawan disajikan pada Tabel 9.

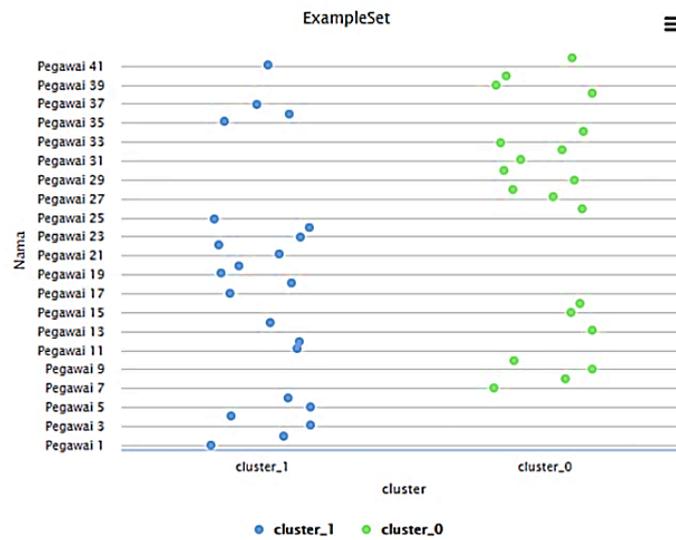
Tabel 9. Hasil pengelompokan.

Nama	<i>Chebyshev</i>		<i>Cluster</i>	
	C1	C2	C1	C2
Pegawai 1	20	16		*
Pegawai 2	28	20		*
Pegawai 3	40	28		*
Pegawai 4	20	10		*
Pegawai 5	28	9		*
Pegawai 6	28	8		*
Pegawai 7	12	14	*	
Pegawai 8	8	16	*	
:	:	:		
Pegawai 35	20	10		*
Pegawai 36	20	16		*
Pegawai 37	28	8		*
Pegawai 38	28	28	*	
Pegawai 39	20	34	*	
Pegawai 40	20	20	*	
Pegawai 41	20	0		*
Pegawai 42	0	20	*	

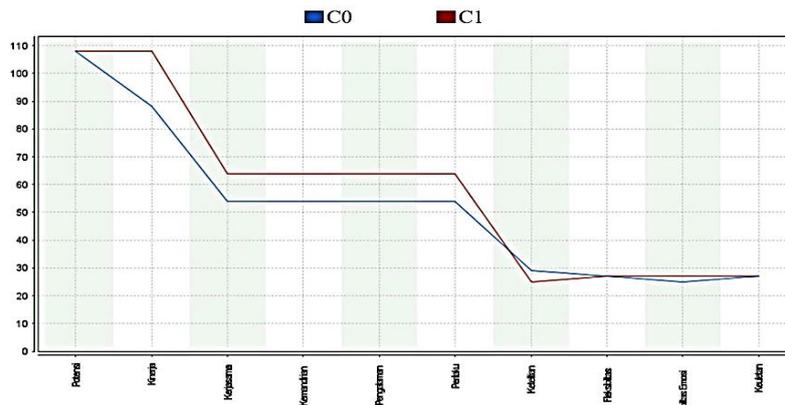
Dari Tabel 9 terdapat 20 data yang termasuk ke dalam *cluster* C1 dan 20 data termasuk ke dalam *cluster* C2. Pemodelan *cluster* menggunakan bantuan *software* Rapidminer disajikan pada Gambar 1. Gambar 1 menunjukkan grafik hasil pemodelan *cluster* menggunakan 42 data, dimana data pada C0 berwarna hijau sebanyak 20 data dan C1 berwarna biru sebanyak 22 data. Dalam menentukan *cluster* mana yang termasuk dalam *cluster* dengan kategori layak dan tidak layak maka dapat dilihat melalui grafik pengelompokan pada Gambar 2.

Berdasarkan Gambar 2, warna merah adalah C1 dan warna biru adalah C0. Dilihat dari grafik tersebut C1 berada pada posisi tertinggi dari C0 sehingga dapat diambil kesimpulan bahwa C1 merupakan *cluster* dengan kategori layak, sementara itu C0 berada pada posisi bawah sehingga dapat disimpulkan bahwa C0 merupakan kategori tidak layak. Hasil tersebut menunjukkan bahwa C1 pada *software* berkorelasi dengan C2 pada perhitungan yang merupakan kategori layak, sementara C0 pada *software* sama dengan C1 pada perhitungan yang termasuk ke dalam kategori tidak layak. Pegawai yang termasuk ke dalam kategori tidak layak adalah pegawai 7, pegawai 8, pegawai 9, pegawai 10, pegawai 13, pegawai 15, pegawai 16, pegawai 26, pegawai 27, pegawai 28, pegawai 29, pegawai 30, pegawai 31, pegawai 32, pegawai 33, pegawai 34, pegawai 38, pegawai 39, pegawai 40, dan pegawai 42. Sedangkan pegawai yang layak dipertahankan adalah pegawai 1, pegawai 2, pegawai 3, pegawai 4, pegawai 5, pegawai 6, pegawai 11, pegawai 12,

pegawai 14, pegawai 17, pegawai 18, pegawai 19, pegawai 20, pegawai 21, pegawai 22, pegawai 23, pegawai 24, pegawai 25, pegawai 35, pegawai 36, pegawai 37, dan pegawai 41.



Gambar 1. Hasil pemodelan *K-Medoids*.



Gambar 2. Grafik plot pengelompokan.

Hasil pengelompokan tersebut kemudian dibandingkan dengan data pengelompokan dari Bankaltimtera Kantor Cabang Tanjung Redeb yang membagi kategori pengelompokan pegawai kontrak menjadi direkomendasikan dan tidak direkomendasikan. Dalam hal ini terdapat 1 pegawai kontrak yang tidak direkomendasikan yaitu pegawai 26, sedangkan untuk kategori direkomendasikan terdapat 40 pegawai, sementara itu terdapat 1 pegawai kontrak yang habis masa kontraknya yaitu pegawai 5. Hasil perbandingan tersebut menunjukkan bahwa terdapat perbedaan jumlah kategori dan hasil pengelompokan, dimana penelitian ini menggunakan dua kategori pengelompokan yang terdiri dari layak dan tidak layak, sedangkan Bankaltimtera menggunakan tiga kategori yang terdiri dari direkomendasikan, tidak direkomendasikan dan habis masa kontrak. Berdasarkan hasil analisis penelitian yang telah dilakukan, pegawai yang berstatus layak perlu direkomendasikan sebagai pegawai yang mendapatkan perpanjangan masa kontraknya. Sedangkan, pegawai berstatus tidak layak, tidak direkomendasikan untuk diperpanjang status keberlanjutan kontraknya.

4. Kesimpulan

Penerapan metode K-Medoids pada penelitian ini untuk mengelompokkan status keberlanjutan karyawan kontrak telah menghasilkan klasterisasi status keberlanjutan karyawan kontrak yang menempatkan 20 Pegawai ke dalam kategori Tidak Layak (C1) dan 22 Pegawai ke dalam kategori Layak (C2). Hasil dari pengujian akurasi *distance measure* menggunakan metode *silhouette coefficient* sebesar 0.3220214 diperoleh bahwa *chebyshev distance* merupakan metode pengukuran jarak yang ideal, dan hasil uji akurasi *cluster* menggunakan metode sum of square (SSE) menunjukkan bahwa nilai *error* terkecil dimiliki oleh 2 *cluster*. Penelitian selanjutnya, diharapkan menggunakan metode uji akurasi, metode pengukuran jarak lainnya dan di komparasi menggunakan metode clustering lain seperti *K-Means*, *Decission Tree* dan sebagainya.

Referensi

- [1] A. Thuda, J. Sari, and A. Maharani, "Employees Perception of Human Capital Practices, Employee's Productivity, and Company Performance," *Integr. J. Bus. Econ.*, vol. 3, no. 3, p. 240, 2019, doi: 10.33019/ijbe.v3i3.188.
- [2] J. Rubery, D. Grimshaw, A. Keizer, and M. Johnson, "Challenges and Contradictions in the 'Normalising' of Precarious Work," *Work. Employ. Soc.*, vol. 32, no. 3, pp. 509–527, 2018, doi: 10.1177/0950017017751790.
- [3] Y. Pomantow, S. L. Mandey, and R. N. Taroreh, "Perbedaan Kinerja Karyawan Tetap Dan Kinerja Karyawan Tidak Tetap pada Perusahaan Daerah Air Minum (PDAM) Kab. Minahasa Tenggara," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 7, no. 4, pp. 5544–5551, 2019.
- [4] E. Lyons, "The Impact of Job Training on Temporary Worker Performance: Field Experimental Evidence from Insurance Sales Agents," *J. Econ. Manag. Strateg.*, vol. 29, no. 1, pp. 122–146, 2020, doi: 10.1111/jems.12333.
- [5] M. Susanti and Jefa, "Prediksi Pengangkatan Karyawan Kontrak Menjadi Karyawan Tetap Menggunakan Decision Tree Pada PT . Baskara Cipta Putra," *J. Bianglala Inform.*, vol. 6, no. 1, pp. 1–7, 2018.
- [6] T. Winarti, D. Priyanto, V. Vydia, and H. Indriyawati, "Penerapan Data Mining Menggunakan Algoritma Naive Bayes ntuk Klasifikasi Perpanjangan Kontrak Kerja Karyawan," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2020.
- [7] T. Boongoen and N. Iam-On, "Cluster Ensembles: A Survey of Approaches with Recent Extensions and Applications," *Comput. Sci. Rev.*, vol. 28, pp. 1–25, 2018, doi: 10.1016/j.cosrev.2018.01.003.
- [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. Artificial Intell. Data Min.*, vol. 2, no. 2, pp. 107–115, 2019.
- [9] K. Annisa, B. S. Ginting, and M. A. Syari, "Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat," *J. Ilmu Komput. dan Inform.*, vol. 06, 2022, doi: <http://dx.doi.org/10.30829/algoritma.v6i1.11624>.
- [10] N. Y. Aung, A. C. Mon, and S. Z. Hlaing, "Performance Analysis of Parallel Clustering on Spark Computing Platform," *2nd Int. Conf. Adv. Inf. Technol.*, vol. 1, no. 2, pp. 39–44, 2018.
- [11] J. A. Talingdan, "Data Mining using Clustering Algorithm as Tool for Poverty Analysis," *ACM Int. Conf. Proceeding Ser.*, pp. 56–59, 2019, doi: 10.1145/3316615.3316672.
- [12] S. Theresia, Suhada, I. S. Saragih, I. S. Damanik, and D. Suhendro, "Pengklastran Gaji Karyawan Pada Pt . Erba Primas Bogor," vol. 4, pp. 395–402, 2020, doi: 10.30865/komik.v4i1.2852.

- [13] L. Purba, S. Saifullah, and R. Dewi, "Pengelompokan Kasus Penyakit Aids Berdasarkan Provinsi Dengan Data Mining K-Medoids Clustering," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 687–694, 2019, doi: 10.30865/komik.v3i1.1679.
- [14] Martanto, S. Anwar, C. L. Rohmat, F. M. Basysyar, and Y. A. Wijaya, "Clustering of Internet Network Usage using the K-Medoid Method," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, pp. 1–6, 2021, doi: 10.1088/1757-899x/1088/1/012036.
- [15] R. N. H. Hutasuhut, H. Okprana, and B. E. Damanik, "Penerapan Data Mining Untuk Menentukan Penerima Program Bidikmisi Menggunakan Algoritma K-Medoids," *TIN Terap. Inform. Nasant.*, vol. 2, no. 11, pp. 667–672, 2022, doi: 10.47065/tin.v2i11.1516.
- [16] N. I. Asriny *et al.*, "Comparison of K-Medoids and Self Organizing Maps Algorithm in Grouping Hydrometeorological Natural Disasters in Java Island," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1077, no. 1, pp. 1–13, 2021, doi: 10.1088/1757-899X/1077/1/012008.
- [17] M. Orisa and A. Faisol, "Analisis Algoritma Partitioning Around Medoid untuk Penentuan Klasterisasi," *J. Teknol. Inf. dan Terap.*, vol. 8, no. 2, pp. 86–90, 2021.
- [18] H. Zayuka, S. M. Nasution, and Y. Purwanto, "Perancangan Dan Analisis Clustering Data Menggunakan Metode K-Medoids Untuk Berita Berbahasa Inggris Design and Analysis of Data Clustering Using K-Medoids Method For English News," *e-Proceeding Eng.*, vol. 4, no. 2, pp. 2182–2190, 2017.
- [19] D. ayu Ilfiana, "Pengklastran Puskesmas di Kabupaten Kudus Menggunakan Metode K-Means dengan Perbandingan Jarak Euclidean dan Chebyshev," *Prism. Pros. Semin. Nas. Mat.* 5, vol. 5, pp. 787–798, 2022.
- [20] N. Puspitasari, H. Haviluddin, and F. U. J. H. Puadi, "Klasterisasi Wilayah Penghasil Tanaman Lada Menggunakan Algoritma K-Means," *Indones. J. Comput. Sci.*, vol. 11, no. 3, 2022.
- [21] R. I. Fajriah, H. Sutisna, and B. K. Simpony, "Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 4, no. 1, pp. 36–49, 2019.
- [22] B. L. Pailan, H. Haviluddin, M. Wati, N. Puspitasari, and E. Budiman, "Analisa Kebutuhan Tenaga Kesehatan Menggunakan Algoritma K-Means," *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 1–9, 2023.
- [23] G. P. I. Rani, A. Aziz, and M. P. T. Sulistyono, "Implementasi Euclidean Dan Chebyshev Distance Pada K-Medoids Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, 2022.
- [24] R. Y. Prabowo, Rahmadwati, and P. Mudjirahardjo, "Klasifikasi Kandungan Nitrogen berdasarkan Warna Daun melalui Color Clustering menggunakan Metode Fuzzy C Means dan Hybrid PSO K-Means," *J. EECCIS*, vol. 12, no. 1, pp. 1–8, 2018.
- [25] E. M. S. Rochman, Miswanto, and H. Suprajitno, "Comparison of Clustering in Tuberculosis Using Fuzzy C-Means and K-Means Methods," *Commun. Math. Biol. Neurosci.*, vol. 41, pp. 1–20, 2022, doi: 10.28919/cmbn/7335.
- [26] N. Puspitasari, A. A. Maulana, and F. Alameka, "K-Means untuk Klasterisasi Daerah Rawan Penyakit Demam Berdarah K-Means for Clustering of Dengue Fever Prone Areas," *J. SISFOTENIKA*, vol. 13, no. 1, pp. 40–52, 2023.
- [27] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, pp. 918–926, 2019, doi: 10.30645/senaris.v1i0.100.
- [28] M. Wati, M. S. Noorlah, A. Tejawati, A. Septiarini, M. Jamil, and N. Puspitasari, "Implementation of The K-means Clustering for The Public Health Center Data," in *2022 International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT)*, 2022, pp. 65–69.