

## Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means

Ibnu Amri Thaher<sup>1</sup>, Anindita Septiariani<sup>2\*</sup>, Novianti Puspitasari<sup>3</sup>

<sup>1,2,3</sup>)Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman  
Jl. Sambaliung No. 9, Samarinda, Indonesia

\*email: anindita@unmul.ac.id

(Naskah masuk: 3 September 2021; diterima untuk diterbitkan: 2 Februari 2022)

**ABSTRAK** – Penelitian ini dilakukan pada Badan Pusat Statistika Kota Samarinda untuk mengelompokkan kualitas kinerja pegawai yang hasilnya dapat dimanfaatkan untuk mengetahui kelompok (cluster) dari setiap pegawainya. Penilaian kualitas kinerja pegawai tersebut dilakukan setiap tahunnya. Hasil penilaian tersebut dikelompokkan menjadi beberapa cluster sebagai bahan pertimbangan bagi pemimpin untuk memberikan rekomendasi jabatan kepada pegawainya. Pengelompokan tersebut dilakukan dengan menggunakan metode k-means. Pengelompokan data dilakukan berdasarkan jarak terdekat ke pusat cluster. Pada penelitian ini diterapkan tiga algoritma untuk menentukan jarak ke pusat centroid untuk melihat perbedaan hasilnya. Sampel data yang digunakan sebanyak 25 pegawai dengan lima parameter yang terdiri dari profesional, integritas, amanah, capaian kinerja pegawai, dan absensi. Data dibuat menjadi 3 cluster yang merupakan jumlah cluster paling optimal berdasarkan hasil pengujian Sum of Square dengan nilai sebesar 3,55 yang merupakan nilai dengan selisih terbesar. Hasil penerapan metode k-means diperoleh sebanyak 12 karyawan berada pada cluster satu, 10 karyawan berada pada cluster kedua, dan 3 karyawan berada pada cluster ketiga. Berdasarkan nilai pusat centroid pada iterasi terakhir disimpulkan bahwa pegawai pada cluster pertama merupakan pegawai yang memiliki nilai terbaik, cluster kedua dan ketiga memiliki nilai yang lebih rendah secara berurutan.

**Kata Kunci** – Clustering; K-Means; Euclidean Distance; Manhattan Distance; Minkowski Distance; Sum of Square.

## Grouping the Quality of Employee Performance for Position Recommendations Using the K-Means Clustering Method

**ABSTRACT** – This study was conducted to classify the quality of employee performance, the results of which can be used by the Central Bureau of Statistics in Samarinda City to determine the clusters of each employee. The quality evaluation of the employee's performance is carried out annually. The assessment results are grouped into several clusters as consideration for leaders to provide job recommendations to their employees. The clustering was applied using the k-means method. Data clustering was performed based on the closest distance to the center of the cluster. In this study, three algorithms were applied to determine the distance to the center of the centroid to see the result difference. The sample data used were 25 employees with five attributes: professionalism, integrity, trustworthiness, employee performance achievements, and absenteeism. The data were clustered into 3 clusters, which are the most optimal number of clusters based on the Sum of Square test results with a value of 3.55, which is the value with the enormous difference. The results of the implementation of the clustering method were obtained: 12 employees are in cluster one, 10 employees are in the second cluster, and 3 employees are in the third cluster. Based on the value of the centroid center in the last iteration, it was concluded that the employees in the first cluster were the employees with the best scores, the second and third clusters tend to have the lower score, respectively.

**Keywords** - Clustering; K-Means; Euclidean Distance; Manhattan Distance; Minkowski Distance; Sum of Square.

### 1. PENDAHULUAN

Sumber daya manusia dari perusahaan berpengaruh terhadap banyak aspek yang dapat

menentukan keberhasilan kinerja perusahaan tersebut. Salah satu proses penting dalam Human Resources Department (HRD) suatu perusahaan atau

badan usaha yaitu proses promosi kenaikan jabatan. Meningkatnya persaingan antar pegawai membuat keputusan promosi menjadi lebih sulit, terutama jika ada beberapa pegawai dengan kemampuan yang sama. Keputusan yang tidak tepat dapat berdampak negatif bagi perusahaan. Hal ini dapat berpengaruh terhadap tingkat kinerja pegawai pada perusahaan tersebut [1].

Promosi kenaikan jabatan membutuhkan suatu seleksi terkait penilaian kualitas serta kinerja terhadap setiap pegawai pada sebuah perusahaan atau instansi pemerintahan. Penilaian kinerja adalah proses evaluasi produktifitas pegawai yang menunjukkan seberapa baik atau buruknya kinerja seorang pegawai dalam pekerjaannya [2]. Selain itu, penilaian kinerja dapat digunakan untuk membantu pimpinan dalam mengantisipasi dan mencegah ketidakpuasan pegawai [3].

Proses penilaian kinerja pegawai yang dilakukan pada perusahaan atau instansi pemerintahan kadang masih bersifat subjektif. Hal itu dapat menyebabkan ketidaksesuaian dalam pelaksanaan penilaian tingkat kualitas kinerja pegawai sehingga dimungkinkan dapat menurunkan kualitas sebuah perusahaan atau instansi pemerintahan. Selama ini masih terdapat instansi pemerintah yang belum memiliki pegawai yang cukup kompeten, hal ini dapat dilihat dari rendahnya produktivitas pegawai di lingkungan instansi pemerintah dan hasil pengukuran kinerja pegawai masih belum dimanfaatkan dengan sesuai [4].

Berdasarkan hal tersebut perlu diadakan sebuah evaluasi serta penilaian kualitas kinerja seorang pegawai berdasarkan fakta dan data serta kualitas dan kinerja tanpa dipengaruhi untuk meningkatkan kualitas sebuah perusahaan maupun instansi pemerintahan. Penilaian kinerja pegawai merupakan bagian yang sangat penting, karena dengan kualitas pegawai yang unggul maka diperoleh hasil kerja yang baik [5]. Selain itu, tujuan perusahaan tercapai, kelangsungan perusahaan dan kaderisasi pegawai dapat terlaksana dengan baik.

Badan Pusat Statistika (BPS) menganggap bahwa penilaian kualitas kinerja pegawai merupakan salah satu hal yang penting. BPS merupakan salah satu lembaga pemerintahan non-Departemen yang bertanggung jawab langsung kepada presiden. BPS melaksanakan tugas pemerintahan bidang statistik sesuai peraturan perundang-undangan. BPS memiliki fungsi pengkajian, penyusunan, dan perumusan di bidang statistika serta hal lainnya yang berkaitan dengan kegiatan statistik pemerintahan.

Permasalahan yang ada pada saat ini dapat diselesaikan menggunakan sistem komputer dengan menerapkan salah satu teknik data mining yaitu *clustering*. *Clustering* merupakan suatu proses pengelompokan data/objek ke dalam kelompok

(*cluster*) tertentu, di mana objek yang ada dalam suatu *cluster* memiliki kesamaan yang tinggi dibanding dengan satu sama lain, tetapi berbeda dengan objek dalam *cluster* lainnya [6]. *Clustering* merupakan *unsupervised learning*, di mana metode yang umum digunakan adalah k-means karena memiliki algoritma yang tergolong sederhana untuk diterapkan, relatif cepat, dan mudah diadaptasi [7]. K-means memiliki kemampuan untuk mengelompokkan sejumlah data dengan waktu komputasi yang relatif cepat dan efisien [8].

Penerapan k-means untuk melakukan pengelompokan data juga sudah banyak diterapkan pada penelitian sebelumnya. Contohnya antara lain untuk penentuan kualitas kinerja pegawai [4], [7], [9], penentuan mahasiswa berprestasi berbasis kelas akademik [10], pengelompokan pemukiman kumuh di kota Bengkulu [11] dan untuk mengevaluasi hasil belajar siswa sekolah menengah kejuruan [12].

Penentuan kualitas kinerja penunjang di Universitas Udayana dengan menggunakan 544 sample data dengan parameter NIP, nama pegawai, status, kinerja NCP, poin hitung kinerja, poin hitung dukungan, poin dukungan, performa poin total, performa titik pengguna. Hasil pada penelitian tersebut yaitu 201 anggota dengan kinerja sangat baik dengan akurasi 37%, 37 anggota dengan kinerja kurang dengan akurasi 7%, 273 anggota dengan kinerja baik dengan akurasi 44%, 64 anggota dengan kinerja buruk dengan akurasi 12%, dan 5 orang dengan performa beragam dengan akurasi 1% [4].

Penilaian kerja aparatur sipil negara di sekretariat DPRD Pemantang Siantar menggunakan sample data sebanyak 23 pegawai. Atribut yang digunakan pada penelitian tersebut yaitu penilaian sasaran kerja pegawai (SKP), dan penilaian perilaku. Hasil dari penelitian tersebut terdapat 1 pegawai di *cluster* satu, 9 pegawai di *cluster* dua, 13 data di *cluster* 3 [7].

Penentuan kualitas kinerja pegawai pada perusahaan bahan kimia menggunakan 38 sample data pegawai dengan parameter nilai kualitas pekerjaan, nilai tanggung jawab, nilai kerjasama, nilai kehadiran, nilai kedisiplinan. Hasil yang diperoleh yaitu 16 data pada cluster 1 dengan akurasi 0,42%, 18 data pada cluster 2 dengan akurasi 0,47%, dan 4 data pada cluster 3 dengan akurasi 0,45% [9].

Pengelompokan mahasiswa berprestasi berbasis akademik menggunakan 10 sample data dengan parameter NIS siswa, nama siswa, nilai tugas, nilai ujian, dan nilai praktek. Hasil penelitian tersebut yaitu sebanyak 4 siswa dengan hasil sangat baik, 2 siswa dengan hasil baik, 2 siswa dengan hasil cukup, dan 2 siswa dengan hasil kurang [10].

Pemetaan pemukiman kumuh di Kota Bengkulu menggunakan 67 sample data kelurahan dengan parameter. Penelitian tersebut menggunakan 7 indikator dan 19 sub indikator. Hasil dari

perhitungan yang dilakukan diperoleh hasil 45 kelurahan masuk dalam *cluster* tidak kumuh, 6 kelurahan kumuh ringan, 1 kelurahan masuk di *cluster* kumuh sedang, dan 15 kelurahan masuk di *cluster* kumuh berat [11].

Evaluasi hasil belajar siswa di sekolah menengah kejuruan dilakukan menggunakan sample data terdiri dari 50 siswa. Analisis dilakukan berdasarkan nilai yang diperoleh siswa kelas X dengan atribut semester 1 sampai 2. Hasil yang diperoleh dari penelitian tersebut yaitu sebanyak 12 iterasi memberikan *output* 20 siswa dengan nilai baik, 22 siswa dengan hasil memuaskan dan akan melakukan pengayaan, dan 8 orang dengan hasil kurang yang harus remedial [12].

Oleh sebab itu, penelitian ini mengusulkan pengelompokan hasil penilaian kualitas kinerja pegawai sebagai bahan pertimbangan pimpinan untuk memberikan rekomendasi jabatan pada Badan Pusat Statistik Kota Samarinda. Pada kasus yang terjadi terdapat sebuah masalah dalam menentukan kelompok pegawai yang dapat menempati posisi jabatan struktural Kasubag. umum. Permasalahan dapat diselesaikan dengan menerapkan *clustering* menggunakan metode k-means yang diukur berdasarkan hasil penilaian dari yang berwenang. Adapun penilaian dilakukan berdasarkan profesionalitas, integrasi, amanah, capaian kinerja pegawai, kehadiran setiap pegawai yang akan dibagi menjadi beberapa *cluster* dan dilakukan pengujian untuk menentukan jumlah *cluster* terbaik.

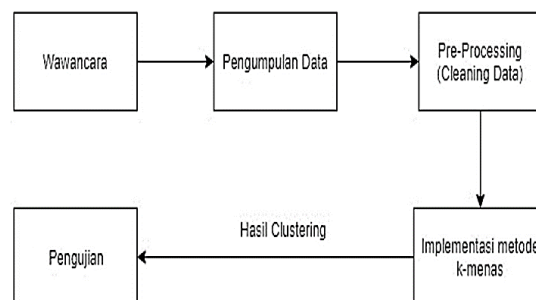
## 2. METODE DAN BAHAN

### 2.1 Tahapan Penelitian

Pada pelaksanaannya, penelitian ini melakukan beberapa tahapan proses, yang terdiri dari wawancara, pengumpulan data, implementasi algoritma, pengujian algoritma. Alur tahapan proses penelitian dapat dilihat pada Gambar 1. Adapun penjelasan setiap tahap adalah sebagai berikut.

- 1) Wawancara: dilakukan untuk mengetahui atribut apa saja yang digunakan sebagai parameter untuk menentukan kualitas kinerja pegawai serta mencari informasi terkait jumlah kelompok data.
- 2) Pengumpulan data: dilakukan pengumpulan data di BPS Kota Samarinda, data yang dikumpulkan yaitu berupa data jumlah pegawai serta data hasil penilaian tentang profesionalitas, integrasi, amanah, capaian kerja pegawai, dan kehadiran dari setiap pegawai yang ada.
- 3) *Pre-Processing (Cleaning data)*: dilakukan proses pembersihan data untuk menghindari duplikasi, inkonsistensi serta data yang tidak sesuai dengan ketentuan dan pembobotan

data berdasarkan ketentuan dengan bobot profesional 30%, Integritas 20%, Amanah 20%, CKP 20%, dan Absensi 20%.



Gambar 1. Tahapan Penelitian Pengelompokan Kinerja Pegawai pada BPS Kota Samarinda

4) Implementasi metode k-means: dilakukan penerapan metode k-means menggunakan data-data yang telah dikumpulkan pada tahap sebelumnya, Implementasi metode dilakukan sesuai dengan ketentuan pada tahap proses wawancara.

5) Pengujian: dilakukan pengujian algoritma dari hasil perhitungan yang telah dilakukan, pengujian algoritma bertujuan untuk menguji seberapa tepat dalam pengelompokan data yang dilakukan atau melakukan pengujian jumlah *cluster* optimal.

### 2.2 K-Means

K-Means merupakan salah satu metode non-hirarki *clustering* yang mencoba membagi data yang ada menjadi satu atau lebih *cluster*. Algoritma mengelompokkan data ke dalam suatu *cluster* berdasarkan kemiripan atau kesamaan karakteristik data sedangkan data dengan karakteristik yang tidak mirip atau berbeda dikelompokkan ke dalam *cluster* yang lain [7]. Sementara itu, k-Means merupakan salah satu metode pengelompokan non-hirarki yang bertujuan mengelompokkan objek sedemikian sehingga jarak-jarak tiap objek ke pusat kelompok dalam satu kelompok adalah minimum [11]. Adapun algoritma dalam metode k-means *clustering* adalah sebagai berikut:

1. Menentukan jumlah k (*cluster*) yang menjadi acuan pengelompokan data.
2. Menentukan titik pusat (*centroid*) dari masing-masing *cluster* yang ada.
3. Menghitung jarak terdekat masing-masing data ke pusat *centroid*, dengan menggunakan salah satu persamaan jarak untuk menentukan jarak terdekat ke pusat *cluster* dengan persamaan jarak *Euclidean*, *Manhattan*, dan *Minkowski*.
4. Menentukan nilai *centroid* baru pada setiap *cluster* yang ada, dengan menghitung rata-rata (*means*) dari data pada tiap *cluster* dengan rumus persamaan (1).

$$Cluster\ Center = \sum \frac{a_i}{n} \dots\dots\dots(1)$$

5. Mengulangi Langkah 3 sampai 5 hingga tidak ada data record yang berpindah *cluster* atau konvergen.

### 2.3 Euclidean Distance

Euclidean *distance* merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari dua buah titik dalam *Euclidean space* (meliputi bidang dua dimensi, tiga dimensi, atau bahkan lebih. Pengukuran jarak dilakukan dengan rumus *Euclidean distance* pada persamaan (2) sebagai berikut [13]:

$$d(x_1, x_2) = ||x_1 - x_2|| = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \dots (2)$$

di mana  $x_1$  adalah data latih dan  $x_2$  adalah data uji

### 2.4 Manhattan Distance

Manhattan *distance* digunakan untuk menghitung perbedaan nilai absolut (mutlak) antara koordinat sepasang objek. Rumus persamaan yang dapat digunakan untuk menghitung jarak dengan Manhattan *distance* dapat dilihat pada persamaan (3) sebagai berikut [14]:

$$d_{\text{Manhattan}}(x, y) = \sum_i^n |x_i - y_i| \dots \dots \dots (3)$$

Keterangan:

$d$  = jarak antara data dengan centroid

$i$  = data ke- $i$

$n$  = jumlah data

$x_i$  = data pada setiap data ke  $i$

$y_i$  = data pada pusat *cluster* ke  $i$

### 2.5 Minkowski Distance

*Minkowski distance* merupakan sebuah metrik dalam ruang vektor dimana suatu norma didefinisikan (*normed vector space*) sekaligus dianggap sebagai generalisasi dari *Euclidean distance* dan *Manhattan distance* [15]. Dalam pengukuran jarak objek menggunakan *minkowski distance* umumnya digunakan nilai  $p$  adalah 1 atau 2. Berikut merupakan rumus persamaan yang dapat digunakan untuk menghitung jarak dapat dilihat pada persamaan (4) sebagai berikut:

$$d_{\text{Minkowski}}(x, y) = (\sum_i^n |x_i - y_i|^p)^{1/p} \dots \dots \dots (4)$$

Keterangan:

$d$  = jarak antara data dengan centroid

$i$  = data ke- $i$

$n$  = jumlah data

$x_i$  = data pada setiap data ke  $i$

$y_i$  = data pada pusat *cluster* ke  $i$

$p$  = power.

### 2.5 Sumber Data

Data yang digunakan pada penelitian ini adalah *data record* pegawai BPS Kota Samarinda dari bulan Januari sampai Desember tahun 2020. Data yang

diperoleh berupa data pegawai yang berjumlah 26 orang. Data setiap *record* terdiri dari hasil penilaian dari yang berwenang tentang sikap profesionalitas (A2), integritas (A3), amanah (A4) yang diambil dari nilai akuntabilitas yang merupakan hasil rata-rata dari ketiga parameter pengukuran tersebut. Parameter selanjutnya adalah nilai rata-rata capaian kinerja pegawai (A5) tahun 2020 yang dilaporkan setiap bulannya oleh setiap karyawan yang terdiri dari Capaian kerja pegawai-Target (CKP-T) dan Capaian kerja pegawai-Realisasi (CKP-R). CKP-T adalah berisi segala kegiatan yang menjadi target pegawai setiap bulannya dan CKP-R adalah capaian kerja pegawai yang terealisasi setiap bulannya berisi hasil penilaian dari yang berwenang tentang kuantitas dan kualitas setiap target kerja yang direalisasikan oleh pegawai setiap bulannya, parameter terakhir adalah rata-rata absensi (A6) kehadiran pegawai BPS Kota Samarinda setiap tahunnya. adapun data pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Sumber Data

No	Nama	A2	A3	A4	A5	A6
1	Peg. 1	98	98	98	98.18	99
2	Peg. 2	98	98	98	99.44	98.67
3	Peg. 3	99	99	99	99.29	99
4	Peg. 4	98	98	98	99.54	98.67
5	Peg. 5	95	95	95	99.16	99
6	Peg. 6	95	95	95	99.17	99
7	Peg. 7	99	99	99	99.27	99
8	Peg. 8	97	97	97	99.50	99
9	Peg. 9	98	98	98	99.12	99
10	Peg. 10	Cuti	Cuti	Cuti	99.25	99
11	Peg. 11	98	98	98	99.50	99
12	Peg. 12	99	99	99	99.13	99
13	Peg. 13	90	90	90	99.02	98.33
14	Peg. 14	98	98	98	99.10	99
15	Peg. 15	99	99	99	99.39	99
16	Peg. 16	99	99	99	99.42	99
17	Peg. 17	98	98	98	99.82	99
18	Peg. 18	98	98	98	99.23	99
19	Peg. 19	99	99	99	99.38	98.67
20	Peg. 20	99	99	99	99.32	99
21	Peg. 21	98	98	98	99.04	99
22	Peg. 22	99	99	99	99.35	99
23	Peg. 23	98	98	98	99.33	99
24	Peg. 24	98	98	98	99.27	99
25	Peg. 25	99	99	99	99.36	99
26	Peg. 26	99	99	99	99.39	99

## 3. HASIL DAN PEMBAHASAN

### 3.1 Cleaning Data

Penelitian ini menggunakan data dari jumlah

seluruh pegawai yang ada di Badan Pusat Statistika Kota Samarinda yang terdiri dari 26 orang karyawan yang kemudian memasuki tahapan proses *cleaning* data di mana ditemukan sebuah data yang tidak konsisten dengan parameter pengelompokan yang telah ditentukan sehingga terjadi reduksi data menjadi 25 orang pegawai. Proses dari *cleaning* data menghapus data pegawai ke-10 dikarenakan tidak memiliki nilai dari atribut *professional*, *integritas*, dan *amanah*. Hasil dari *cleaning* data juga melalui proses pembobotan sesuai dengan ketentuan dari BPS kota Samarinda. Adapun hasil *cleaning* data dan pembobotan dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Cleaning* Dan Pembobotan

No	Nama	A2	A3	A4	A5	A6
1	Peg. 1	29.4	19.6	19.6	19.64	9.9
2	Peg. 2	29.4	19.6	19.6	19.89	9.87
3	Peg. 3	29.7	19.8	19.8	19.86	9.9
4	Peg. 4	29.4	19.6	19.6	19.91	9.87
5	Peg. 5	28.5	19	19	19.83	9.9
6	Peg. 6	28.5	19	19	19.83	9.9
7	Peg. 7	29.7	19.8	19.8	19.85	9.9
8	Peg. 8	29.1	19.4	19.4	19.9	9.9
9	Peg. 9	29.4	19.6	19.6	19.82	9.9
10	Peg. 11	29.4	19.6	19.6	19.9	9.9
11	Peg. 12	29.7	19.8	19.8	19.83	9.9
12	Peg. 13	27	18	18	19.80	9.03
13	Peg. 14	29.4	19.6	19.6	19.82	9.9
14	Peg. 15	29.7	19.8	19.8	19.88	9.9
15	Peg. 16	29.7	19.8	19.8	19.88	9.9
16	Peg. 17	29.4	19.6	19.6	19.96	9.9
17	Peg. 18	29.4	19.6	19.6	19.85	9.9
18	Peg. 19	29.7	19.8	19.8	19.88	9.87
19	Peg. 20	29.7	19.8	19.8	19.86	9.9
20	Peg. 21	29.7	19.8	19.8	19.87	9.9
21	Peg. 22	29.4	19.6	19.6	19.87	9.9
22	Peg. 23	29.4	19.6	19.6	19.85	9.9
23	Peg. 24	29.7	19.8	19.8	19.87	9.9
24	Peg. 25	29.7	19.8	19.8	19.88	9.9
25	Peg. 26	29.4	19.6	19.6	19.81	9.9

### 3.2 Implementasi K-Means

Pada tahapan ini dilakukan penerapan proses perhitungan algoritma *k-means* pada data yang telah melewati proses *cleaning* data adapun tahapan yang terjadi dimulai dari menentukan nilai *k*, menentukan nilai random sebagai pusat *centroid*, menghitung jarak terdekat dari pusat *cluster*, menentukan pusat *centroid* baru, melakukan perulangan jika hasil data berpindah *cluster*.

#### 1. Penentuan Nilai K

Penentuan nilai *k* pada proses algoritma dilakukan untuk menentukan jumlah *cluster* atau pengelompokan data yang akan terjadi. Pada

penelitian ini jumlah *cluster* yang ditentukan dimulai dari jumlah *k*=2 sampai dengan *k*=4 yang kemudian akan melewati proses perhitungan algoritma *k-means* berdasarkan jumlah *k* telah yang ditentukan kemudian dari hasil pengelompokan data dari setiap nilai *k* akan dilakukan pengujian jumlah *k* yang paling optimal menggunakan teknik *Sum of Square* dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* (*k*) yang akan membentuk siku pada suatu titik.

#### 2. Penentuan Centroid Awal

Penentuan nilai *centroid* awal menggunakan nilai random untuk menentukan titik pusat *centroid* pada awal iterasi. Nilai random yang didapat adalah pada data ke 8, 22, dan 4 sebagai pusat *centroid* awal C1, C2, dan C3. Rincian data *centroid* awal dapat dilihat pada Tabel 3, yang terdiri dari *centroid* (A1), *professional* (A2), *integritas* (A3), *amanah* (A4), CPK (A5), dan *absensi* (A6).

Tabel 3. Pusat Centroid Awal

A1	A2	A3	A4	A5	A6
C1	29.4	19.6	19.6	19.824	9.9
C2	29.7	19.8	19.8	19.872	9.9
C3	28.5	19	19	19.832	9.9

#### 3. Menghitung Jarak Terdekat Ke Pusat Cluster

Perhitungan jarak terdekat ke pusat *cluster* menggunakan tiga persamaan untuk melakukan analisis perbandingan antara ketiganya. Hasil analisis dari perbandingan ketiganya didapatkan pengelompokan data ke dalam *cluster* diperoleh *cluster* yang sama dengan hasil perhitungan yang berbeda oleh setiap metode perhitungan jarak.

Adapun proses perhitungan yang digunakan untuk melakukan perhitungan jarak ke pusat *cluster* dengan menggunakan persamaan (2.2), (2.3), (2.4) adalah sebagai berikut:

##### a. Euclidean Distance

Pengukuran yang digunakan untuk menentukan jarak terdekat ke pusat *cluster* dengan menggunakan persamaan *Euclidean distance* berdasarkan data pada Tabel 2 adalah sebagai berikut:

- 1) Perhitungan data ke-1 menuju pusat *cluster* 1 dengan persamaan *Euclidean*

$$d_{(1,1)} = \sqrt{(29,4 - 29,4)^2 + (19,6 - 19,6)^2 + \dots + (9,9 - 9,9)^2}$$

$$d_{(1,1)} = \sqrt{(0)^2 + (0)^2 + (0)^2 + (0,188)^2 + (0)^2}$$

$$d_{(1,1)} = 0,188$$

- 2) Perhitungan data ke-1 menuju pusat *cluster* 2 dengan persamaan *Euclidean*

$$d_{(1,2)} = \sqrt{(29,4 - 29,7)^2 + (19,6 - 19,8)^2 + \dots + (9,9 - 9,9)^2}$$

$$d_{(1,2)} = \sqrt{(-0,3)^2 + (-0,2)^2 + (-0,2)^2 + (-0,236)^2 + (0)^2}$$

$$d_{(1,2)} = \sqrt{0,09 + 0,04 + 0,04 + 0,055696}$$

$$d_{(1,2)} = 0,475$$

- 3) Perhitungan data ke-1 menuju pusat *cluster* 3 dengan persamaan *Euclidean*

$$d_{(1,3)} = \sqrt{(29,4 - 28,5)^2 + (19,6 - 19)^2 + \dots + (9,9 - 9,9)^2}$$

$$d_{(1,3)} = \sqrt{(0,9)^2 + (0,6)^2 + (-0,6)^2 + (-0,196)^2 + (0)^2}$$

$$d_{(1,3)} = \sqrt{0,81 + 0,36 + 0,36 + 0,038416}$$

$$d_{(1,3)} = 1,252$$

**b. Manhattan Distance**

Pengukuran yang digunakan untuk menentukan jarak terdekat ke pusat cluster dengan menggunakan persamaan *Manhattan distance* berdasarkan data pada Tabel 2 adalah sebagai berikut:

- 1) Perhitungan data ke-1 menuju pusat cluster 1 dengan persamaan *Manhattan*

$$d_{(1,1)} = |29,4 - 29,4| + |19,6 - 19,6| + \dots + |9,9 - 9,9|$$

$$d_{(1,1)} = (0) + (0) + (0) + (0,188) + (0)$$

$$d_{(1,1)} = 0,188$$

- 2) Perhitungan data ke-1 menuju pusat cluster 2 dengan persamaan *Manhattan*

$$d_{(1,2)} = |29,4 - 29,7| + |19,6 - 19,8| + \dots + |9,9 - 9,9|$$

$$d_{(1,2)} = (0,3) + (0,2) + (0,2) + (0,236) + (0)$$

$$d_{(1,2)} = 0,936$$

- 3) Perhitungan data ke-1 menuju pusat cluster 3 dengan persamaan *Manhattan*

$$d_{(1,3)} = |29,4 - 28,5| + |19,6 - 19,0| + \dots + |9,9 - 9,9|$$

$$d_{(1,3)} = (0,9) + (0,6) + (0,6) + (0,196) + (0)$$

$$d_{(1,3)} = 2,296$$

**c. Minkowski Distance**

Pengukuran yang digunakan untuk menentukan jarak terdekat ke pusat cluster dengan menggunakan persamaan *Minkowski Distance* berdasarkan data pada Tabel 2 adalah sebagai berikut:

- 1) Perhitungan data ke-1 menuju pusat cluster 1 dengan persamaan *Minkowski*

$$d_{(1,1)} = ((|29,4 - 29,4|)^2 + (|19,6 - 19,6|)^2 + \dots + (|9,9 - 9,9|)^2)^{1/2}$$

$$d_{(1,1)} = 0,036^{1/2}$$

$$d_{(1,1)} = 0,188$$

- 2) Perhitungan data ke-1 menuju pusat cluster 2 dengan persamaan *Minkowski*

$$d_{(1,2)} = ((|29,4 - 29,7|)^2 + (|19,6 - 19,8|)^2 + \dots + (|9,9 - 9,9|)^2)^{1/2}$$

$$d_{(1,2)} = 0,1764^{1/2}$$

$$d_{(1,2)} = 0,475$$

- 3) Perhitungan data ke-1 menuju pusat cluster 3 dengan persamaan *Minkowski*

$$d_{(1,2)} = ((|29,4 - 28,5|)^2 + (|19,6 - 19,0|)^2 + \dots + (|9,9 - 9,9|)^2)^{1/2}$$

$$d_{(1,2)} = 1,5364^{1/2}$$

$$d_{(1,2)} = 1,252$$

Perhitungan jarak menggunakan *Manhattan*, *Euclidean*, dan *Minkowski distance* untuk menghitung jarak terdekat ke pusat centroid menghasilkan perhitungan jarak yang berbeda tetapi memperoleh hasil pengelompokan yang sama terhadap setiap data. Hasil perhitungan jarak diperlihatkan pada Tabel 4 dan Tabel 5.

**4. Menghitung Pusat Centroid Baru**

Setelah mendapatkan hasil pengelompokan data pada iterasi pertama terhadap setiap cluster selanjutnya menentukan nilai dari pusat centroid baru dengan cara menghitung rata-rata dari setiap total record pada setiap cluster yang ada. Menghitung

kembali data untuk mendapatkan pusat cluster yang baru berdasarkan nilai rata-rata dari masing-masing variabel pada masing-masing cluster.

Tabel 4. Hasil Perhitungan menggunakan *Euclidean Distance* dan *Minkowski Distance*

No	Nama	C1	C2	C3	C1	C2	C3
1	Peg. 1	0.188	0.475	1.252	*		
2	Peg. 2	0.072	0.414	1.239	*		
3	Peg. 3	0.414	0.014	1.649		*	
4	Peg. 4	0.090	0.415	1.240	*		
5	Peg. 5	1.237	1.650	0.000			*
6	Peg. 6	1.237	1.650	0.002			*
7	Peg. 7	0.413	0.018	1.649		*	
8	Peg. 8	0.419	0.825	0.827	*		
9	Peg. 9	0.000	0.415	1.237	*		
10	Peg. 11	0.076	0.413	1.239	*		
11	Peg. 12	0.412	0.046	1.649		*	
12	Peg. 13	3.411	3.811	2.237			*
13	Peg. 14	0.004	0.416	1.237	*		
14	Peg. 15	0.416	0.006	1.650		*	
15	Peg. 16	0.417	0.012	1.650		*	
16	Peg. 17	0.140	0.422	1.244	*		
17	Peg. 18	0.022	0.413	1.237	*		
18	Peg. 19	0.417	0.033	1.650		*	
19	Peg. 20	0.414	0.008	1.650		*	
20	Peg. 21	0.415	0.002	1.650		*	
21	Peg. 22	0.042	0.412	1.237	*		
22	Peg. 23	0.030	0.413	1.237	*		
23	Peg. 24	0.415	0.000	1.650		*	
24	Peg. 25	0.416	0.006	1.650		*	
25	Peg. 26	0.016	0.417	1.237	*		

Adapun perhitungan pada masing - masing cluster sebagai berikut:

- a. *Cluster center profesional pada cluster 1:*  
 $Cluster\ Center = \frac{(29,4 + 29,4 + \dots + 29,4 + 29,4 + 29,4)}{12}$   
 $Cluster\ Center = 29,375$
- b. *Cluster center integritas pada cluster 1*  
 $Cluster\ Center = \frac{(19,6 + 19,6 + \dots + 19,6 + 19,4 + 19,6 + 19,6)}{12}$   
 $Cluster\ Center = 19,583$
- c. *Cluster center amanah pada cluster 1*  
 $Cluster\ Center = \frac{(19,6 + 19,6 + \dots + 19,6 + 19,4 + 19,6 + 19,6)}{12}$   
 $Cluster\ Center = 19,583$
- d. *Cluster center ckp pada cluster 1*  
 $Cluster\ Center = \frac{(19,63 + 19,88 + \dots + 19,86 + 19,85 + 19,8)}{12}$   
 $Cluster\ Center = 19,851$
- e. *Cluster center absensi pada cluster 1*  
 $Cluster\ Center = \frac{(19,63 + 19,88 + \dots + 19,86 + 19,85 + 19,8)}{12}$   
 $Cluster\ Center = 9,895$
- f. *Cluster center profesional pada cluster 2*  
 $Cluster\ Center = \frac{(29,7 + 29,7 + \dots + 29,7 + 29,7 + 29,7)}{10}$   
 $Cluster\ Center = 29,7$
- g. *Cluster center integritas pada cluster 2*  
 $Cluster\ Center = \frac{(19,8 + 19,8 + \dots + 19,8 + 19,8 + 19,8 + 19,8)}{10}$   
 $Cluster\ Center = 19,8$
- h. *Cluster center amanah pada cluster 2*

- $Cluster\ Center = \frac{(19,8 + 19,8 + \dots + 19,8 + 19,8 + 19,8 + 19,8)}{10}$   
 $Cluster\ Center = 19,8$
- i. *Cluster center* ckp pada *cluster 2*  
 $Cluster\ Center = \frac{(19,85 + 19,85 + \dots + 19,86 + 19,87 + 19,87)}{10}$   
 $Cluster\ Center = 19,866$
- j. *Cluster center* absensi pada *cluster 2*  
 $Cluster\ Center = \frac{(9,9 + 9,9 + \dots + 9,86 + 9,9 + 9,9 + 9,9)}{10}$   
 $Cluster\ Center = 9,897$
- k. *Cluster center* professional pada *cluster 3*  
 $Cluster\ Center = \frac{28,5 + 28,5 + 27}{3}$   
 $Cluster\ Center = 28$
- l. *Cluster center* integritas pada *cluster 3*  
 $Cluster\ Center = \frac{19 + 19 + 18}{3}$   
 $Cluster\ Center = 18,667$
- m. *Cluster center* amanah pada *cluster 3*  
 $Cluster\ Center = \frac{19 + 19 + 18}{3}$   
 $Cluster\ Center = 18,667$
- n. *Cluster center* ckp pada *cluster 3*  
 $Cluster\ Center = \frac{19,83 + 19,83 + 19,83}{25}$   
 $Cluster\ Center = 19,823$
- o. *Cluster center* absensi pada *cluster 3*  
 $Cluster\ Center = \frac{19,9 + 19,9 + 9,03}{3}$   
 $Cluster\ Center = 9,611$

Tabel 5. Hasil Perhitungan Menggunakan *Manhattan Distance*

No	Nama	C1	C2	C3	C1	C2	C3
1	Peg. 1	0.188	0.936	2.296	*		
2	Peg. 2	0.031	0.717	2.123	*		
3	Peg. 3	0.734	0.014	2.826		*	
4	Peg. 4	0.051	0.697	2.143	*		
5	Peg. 5	2.092	2.840	0.000			*
6	Peg. 6	2.090	2.838	0.002			*
7	Peg. 7	0.730	0.018	2.822		*	
8	Peg. 8	0.624	1.372	1.468	*		
9	Peg. 9	0.000	0.748	2.092	*		
10	Peg. 11	0.076	0.672	2.168	*		
11	Peg. 12	0.702	0.046	2.794		*	
12	Peg. 13	6.487	7.235	4.395			*
13	Peg. 14	0.004	0.752	2.088	*		
14	Peg. 15	0.754	0.006	2.846		*	
15	Peg. 16	0.760	0.012	2.852		*	
16	Peg. 17	0.140	0.608	2.232	*		
17	Peg. 18	0.022	0.726	2.114	*		
18	Peg. 19	0.719	0.029	2.811		*	
19	Peg. 20	0.740	0.008	2.832		*	
20	Peg. 21	0.746	0.002	2.838		*	
21	Peg. 22	0.042	0.706	2.134	*		
22	Peg. 23	0.030	0.718	2.122	*		
23	Peg. 24	0.748	0.000	2.840		*	
24	Peg. 25	0.754	0.006	2.846		*	
25	Peg. 26	0.016	0.764	2.076	*		

Penentuan pusat centroid dilakukan pada setiap iterasi berikutnya dalam proses perhitungan menggunakan k-means. Penentuan pusat centroid baru dengan cara menghitung rata-rata hasil

penjumlahan data seluruh record yang ada pada cluster tersebut menggunakan persamaan (4). Nilai pusat centroid baru dapat dilihat pada Tabel 6.

Tabel 6. Pusat Centroid Baru

A1	A2	A3	A4	A5	A6
C1	29.375	19.583	19.583	19.851	9.895
C2	29.700	19.800	19.800	19.866	9.897
C3	28.000	18.667	18.667	19.823	9.611

### 5. Perulangan Algoritma dan Hasil

Tahapan akhir dalam proses pengelompokan menggunakan algoritma k-means melakukan perulangan jika data cluster pada iterasi tersebut berbeda dengan iterasi sebelumnya dan akan berhenti dan menghasilkan sebuah pengelompokan data final apabila data cluster pada iterasi tersebut sama dengan iterasi sebelumnya. Pada penelitian ini iterasi pada proses perhitungan melewati algoritma k-means terjadi sebanyak dua kali dan memberikan hasil nilai cluster data yang sama dengan iterasi sebelumnya, sehingga membuat proses pengelompokan data mencapai tahapan final dengan data hasil pengelompokan berupa 12 pegawai pada cluster pertama, 10 pegawai pada cluster kedua, dan 3 pegawai pada cluster ketiga. Adapun hasil dari pengelompokan data pada iterasi terakhir dapat dilihat pada Tabel 7 dan 8.

Tabel 7. Hasil Pengelompokan Data Menggunakan Euclidean Dan Minkowski

No	Nama	C1	C2	C3	C1	C2	C3
1	Peg. 1	0.218	0.472	1.955	*		
2	Peg. 2	0.057	0.414	1.942	*		
3	Peg. 3	0.447	0.009	2.354		*	
4	Peg. 4	0.072	0.416	1.943	*		
5	Peg. 5	1.203	1.650	0.746			*
6	Peg. 6	1.203	1.650	0.746			*
7	Peg. 7	0.447	0.012	2.354		*	
8	Peg. 8	0.381	0.825	1.541	*		
9	Peg. 9	0.044	0.414	1.946	*		
10	Peg. 11	0.060	0.414	1.947	*		
11	Peg. 12	0.447	0.040	2.354		*	
12	Peg. 13	3.376	3.810	1.491			*
13	Peg. 14	0.047	0.415	1.946	*		
14	Peg. 15	0.448	0.012	2.355		*	
15	Peg. 16	0.448	0.018	2.355		*	
16	Peg. 17	0.118	0.424	1.951	*		
17	Peg. 18	0.035	0.413	1.946	*		
18	Peg. 19	0.448	0.031	2.351		*	
19	Peg. 20	0.447	0.004	2.355		*	
20	Peg. 21	0.447	0.005	2.355		*	
21	Peg. 22	0.038	0.412	1.946	*		
22	Peg. 23	0.035	0.412	1.946	*		
23	Peg. 24	0.447	0.007	2.355		*	
24	Peg. 25	0.448	0.012	2.355		*	
25	Peg. 26	0.055	0.416	1.946	*		

Tabel 8. Hasil Pengelompokan Menggunakan Data Manhattan

No	Nama	C1	C2	C3	C1	C2	C3
1	Peg. 1	0.279	0.933	3.743	*		
2	Peg. 2	0.123	0.752	3.587	*		
3	Peg. 3	0.771	0.011	4.290		*	
4	Peg. 4	0.143	0.772	3.607	*		
5	Peg. 5	2.066	2.837	1.464			*
6	Peg. 6	2.064	2.835	1.466			*
7	Peg. 7	0.767	0.015	4.286		*	
8	Peg. 8	0.696	1.437	2.932	*		
9	Peg. 9	0.091	0.745	3.556	*		
10	Peg. 11	0.113	0.737	3.632	*		
11	Peg. 12	0.789	0.043	4.258		*	
12	Peg. 13	6.450	7.226	2.931			*
13	Peg. 14	0.095	0.749	3.559	*		
14	Peg. 15	0.791	0.015	4.310		*	
15	Peg. 16	0.797	0.021	4.316		*	
16	Peg. 17	0.177	0.801	3.696	*		
17	Peg. 18	0.069	0.723	3.578	*		
18	Peg. 19	0.811	0.040	4.275		*	
19	Peg. 20	0.777	0.005	4.296		*	
20	Peg. 21	0.783	0.007	4.302		*	
21	Peg. 22	0.079	0.703	3.598	*		
22	Peg. 23	0.067	0.715	3.586	*		
23	Peg. 24	0.785	0.009	4.304		*	
24	Peg. 25	0.791	0.015	4.310		*	
25	Peg. 26	0.107	0.761	3.571	*		

Pada iterasi terakhir di temukan pusat centroid untuk menentukan label pada pengelompokan data yang dilakukan. Adapun pusat *centroid* pada iterasi dengan *Euclidean distance* terakhir dapat dilihat pada Tabel 9.

Tabel 9. Pusat Centroid Pada Iterasi Terakhir

A1	A2	A3	A4	A5	A6
C1	29.375	19.583	19.583	19.851	9.895
C2	29.700	19.800	19.800	19.866	9.897
C3	28.000	18.667	18.667	19.823	9.611

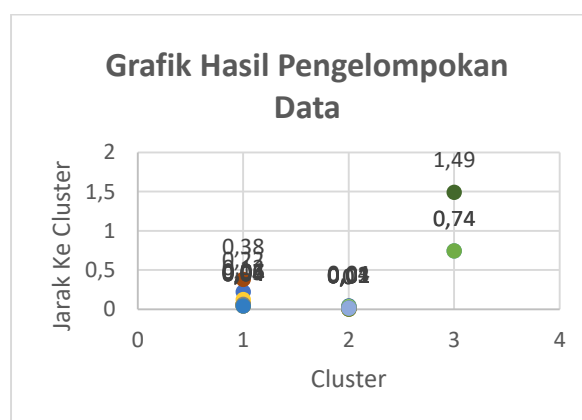
Dari nilai *centroid* akhir yang terbentuk pada iterasi terakhir dapat dilihat bahwasanya pada C2 merupakan centroid yang memiliki nilai-nilai tertinggi dibanding dengan nilai centroid lain. Berdasarkan hal tersebut dapat disimpulkan data yang masuk pada *cluster* tersebut merupakan data-data dengan nilai yang sangat baik, sehingga keterangan yang dapat diberikan pada pengelompokan data yang telah dilakukan dapat dilihat pada Tabel 10.

Proses perhitungan menggunakan algoritma k-means berakhir pada iterasi kedua dengan data record yang tidak berpindah dari iterasi sebelumnya sehingga ditemukan hasil pengelompokan data. Adapun grafik hasil pengelompokan data dapat dilihat pada Gambar 2.

Tabel 10. Label Pengelompokan Data

Cluster	Keterangan
C1	Baik
C2	Sangat Baik
C3	Cukup

Gambar 2 menunjukkan pada grafik yang ada menampilkan kedekatan antara data record pada cluster tersebut dengan data record di sekitarnya dalam satu cluster tetapi mengalami perbedaan jarak antara data pada cluster yang berbeda, sehingga membuktikan metode k-means dapat mengelompokkan data berdasarkan tingkat kedekatan.



Gambar 2. Grafik Hasil Clustering

Disimpulkan bahwa data pada *cluster* kedua merupakan data-data pengelompokan dengan hasil penilaian terbaik berdasarkan kepada nilai pusat centroid pada akhir iterasi dengan jumlah pengelompokan sebanyak tiga *cluster* yang merupakan *cluster* optimal yang dapat digunakan berdasarkan hasil pengujian dengan *sum of square error* untuk menentukan jumlah *cluster* terbaik

### 3.3 Pengujian Algoritma

Pengujian algoritma dilakukan untuk mencari nilai k optimal sehingga diperoleh jumlah cluster terbaik. Pencarian jumlah cluster terbaik digunakan metode Sum of Square untuk menentukan jumlah k optimal. Persamaan yang digunakan sebagai berikut [16]:

$$SST = \sum_{k=1}^k \sum_i^n (x_i - y_i)^2 \dots \dots \dots (5)$$

Pengujian algoritma dengan menggunakan metode Elbow dengan perhitungan *sum of square error* untuk menentukan cluster yang paling optimal. Jumlah cluster yang diuji dimulai dari nilai k=2, k=3, k=4. *Cluster* paling optimal didapatkan jika nilai *Sum of Square Error* pada *cluster* tersebut mengalami penurunan paling besar maka nilai *cluster* tersebut adalah yang terbaik. Pengujian ini dilakukan menggunakan data 25 karyawan yang telah melalui proses algoritma k-means dan memperoleh hasil pengelompokan data menggunakan *Euclidean*



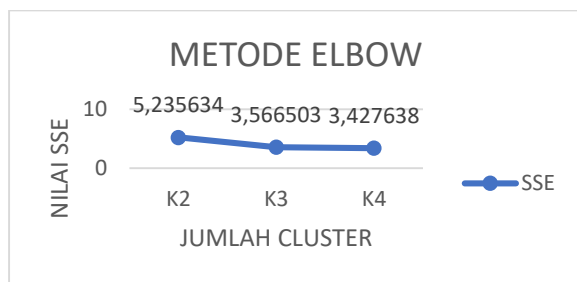
Distance pada iterasi terakhir. Adapun nilai dari dari *sum of square error* pada setiap jumlah k uji dapat dilihat pada Tabel 11.

Tabel 11. Hasil *Sum of Square Error* Tiap Cluster

Cluster	Hasil SSE	Penurunan
K2	5.235634	0
K3	3.566503	1.669131
K4	3.427638	0.138865

Berdasarkan hasil analisis yang diperoleh dari perhitungan pengujian menggunakan persamaan *sum of square error* ditemukan jumlah penurunan paling besar yang terdapat pada k=3 dengan nilai penurunan yaitu 1,669131 sehingga dapat dipastikan cluster paling optimal yang dapat digunakan pada penelitian ini adalah berjumlah 3 cluster.

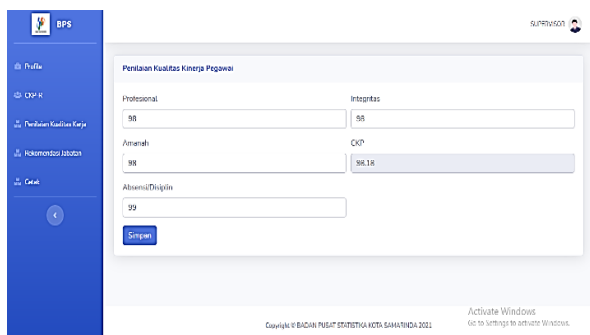
Hasil pengujian nilai k optimal juga dapat dibuktikan dari grafik hasil penggunaan metode elbow yang terbentuk memperlihatkan pembentukan sudut penurunan terbesar pada k=3. Adapun Grafik dari hasil perhitungan nilai SSE dapat dilihat pada Gambar 3.



Gambar 3. Hasil Grafik Metode Elbow

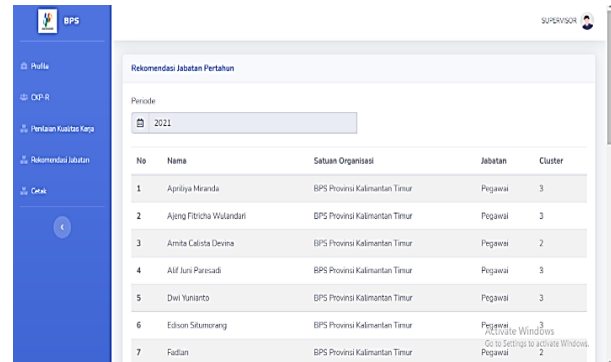
### 3.3 Implementasi Program

Penerapan metode *k-means* untuk menentukan kualitas kinerja pegawai badan pusat statistika kota Samarinda diimplementasikan dalam sebuah sistem berbasis web yang di bangun untuk memudahkan dalam proses perhitungannya adapun tampilan pada sistem yang dibangun dapat dilihat pada Gambar 4.



Gambar 4. Tampilan Halaman Input Data

Adapun hasil dari proses perhitungan *k-means* yang terjadi dapat dilihat pada tampilan sistem yang dibangun pada Gambar 5.



Gambar 5. Tampilan Halaman Hasil Pengelompokan

Sistem yang dibangun untuk memudahkan proses perhitungan penentuan kualitas kinerja pegawai badan pusat statistika kota Samarinda dibangun dengan sistem berbasis website dengan bahasa pemrograman php untuk merancang sistem tersebut. adapun *sources code* yang digunakan adalah sebagai berikut:

#### 1. Menentukan Centeroid Awal

```

$cluster = 3;
$nomorpertanyaan = array();

$ceknomor = array();
for ($i = 0; $i < $cluster; $i++) {
    $nomor = rand(0, $jumlaharray - 1);
    while (in_array($nomor, $ceknomor))
    $ceknomor
    {
        $nomor = rand(0, $jumlaharray);
    }
    $ceknomor[$i] = $nomor;
    $nomorpertanyaan[$i] = $nomor;
}
$centeroid = [0, $nomorpertanyaan[0],
$nomorpertanyaan[1], $nomorpertanyaan[2]];
    
```

#### 2. Menentukan Jarak Terdekat Kepusat Centeroid

```

$datacluster[$i][$j] =
number_format(float)sqrt((pow(($data[$i][&#39;profesional&#39;]; -
esional&#39;]; -
$centeroid[$j][&#39;profesional&#39;];), 2) +
(pow(($data[$i][&#39;integritas&#39;]; -
$centeroid[$j][&#39;integritas&#39;];), 2) +
(pow(($data[$i][&#39;amanah&#39;]; -
$centeroid[$j][&#39;amanah&#39;];), 2) +
(pow(($data[$i][&#39;ckp&#39;]; -
$centeroid[$j][&#39;ckp&#39;];), 2) +
(pow(($data[$i][&#39;absensi&#39;]; -
$centeroid[$j][&#39;absensi&#39;];), 2))), 2);

//Manhattan Distance
//$datacluster[$i][$j] =
number_format(float)($data[$i][&#39;profesional&#39;]; -
($data[$i][&#39;integritas&#39;]; -
    
```

```
//$data[$centeroid[$j]][&#39;integritas&#39;]) +
($data[$i][&#39;amanah&#39;] -
//$data[$centeroid[$j]][&#39;amanah&#39;]) +
($data[$i][&#39;ckp&#39;] -
$data[$centeroid[$j]][&#39;ckp&#39;]) +
//($data[$i][&#39;absensi&#39;] -
$data[$centeroid[$j]][&#39;absensi&#39;]), 2);

//Minkowski Distance
//$datacluster[$i][$j] =
number_format((float)pow(pow(($data[$i][&#39;prof
esional&#39;] -
//$data[$centeroid[$j]][&#39;profesional&#39;]), 2) +
pow(($data[$i][&#39;integritas&#39;] -
//$data[$centeroid[$j]][&#39;integritas&#39;]), 2) +
pow(($data[$i][&#39;amanah&#39;] -
//$data[$centeroid[$j]][&#39;amanah&#39;]), 2) +
pow(($data[$i][&#39;ckp&#39;] -
//$data[$centeroid[$j]][&#39;ckp&#39;]), 2) +
pow(($data[$i][&#39;absensi&#39;] -
//$data[$centeroid[$j]][&#39;absensi&#39;]), 2), 0.5),
2);
$datanya = [];
for ($j = 0; $j &lt; $cluster; $j++) {
# code...
array_push($datanya, $datacluster[$i][$j + 1]);
}
$datapalingkecil = min($datanya);
for ($j = 1; $j &lt;= $cluster; $j++) {
if ($datacluster[$i][$j] == $datapalingkecil) {
$datacluster[$i][$cluster + 2 - 1] = $j;
}}
```

### 3. Menghitung Pusat Centeroid Baru

```
for ($j = 1; $j &lt;= $cluster; $j++)
{
# code...
if ($datacluster[$i][$cluster + 2 - 1] == $j) {
$centeroidbaru[$i][$j][1] =
$data[$i][&#39;profesional&#39;];
$centeroidbaru[$i][$j][2] =
$data[$i][&#39;integritas&#39;];
$centeroidbaru[$i][$j][3] =
$data[$i][&#39;amanah&#39;];
$centeroidbaru[$i][$j][4] = $data[$i][&#39;ckp&#39;];
$centeroidbaru[$i][$j][5] =
$data[$i][&#39;absensi&#39;];
//count
for ($k = 1; $k &lt;= 6; $k++) {
$jumlah[$j][2][$k] += 1;}}
//sum
for ($j = 1; $j &lt;= $cluster; $j++) {
for ($k = 1; $k &lt;= 6; $k++) {
$jumlah[$j][1][$k] += $centeroidbaru[$i][$j][$k];
}}
//avrage
for ($i = 1; $i &lt;= $cluster; $i++) {
for ($j = 1; $j &lt;= 6; $j++) {
```

```
$jumlah[$i][3][$j] = number_format($jumlah[$i][1][$j]
/ $jumlah[$i][2][$j], 2);
}}
```

## 4. KESIMPULAN

Penelitian pengelompokan kualitas kinerja pegawai untuk rekomendasi jabatan pada BPS Kota Samarinda dapat disimpulkan bahwa jumlah *cluster* optimal yang paling baik untuk digunakan adalah  $k=3$  yang mempunyai selisih nilai paling besar dengan nilai SSE 3.566503. Hasil dari pengelompokan terhadap 25 orang karyawan memperoleh hasil 12 orang pegawai di *cluster* baik, 10 orang pegawai di *cluster* sangat baik, dan 3 orang pegawai di *cluster* cukup. Analisis perbandingan dalam perhitungan jarak untuk menentukan jarak terdekat ke pusat centroid dengan menggunakan persamaan Euclidean *distance*, Manhattan *distance*, dan Minkowski *distance* pada penelitian ini diperoleh hasil pengelompokan yang sama.

## DAFTAR PUSTAKA

- [1] F. Frieyadie, "Penerapan Metode Simple Additive Weight (Saw) Dalam Sistem Pendukung Keputusan Promosi Kenaikan Jabatan," *J. Pilar Nusa Mandiri*, vol. 12, no. 1, pp. 37-45, 2016, doi: 10.33480/pilar.v12i1.257.
- [2] M. Martin and Y. Nataliani, "Klasterisasi kinerja karyawan menggunakan algoritma fuzzy c-means," *Aiti*, vol. 17, no. 2, pp. 118-129, 2021, doi: 10.24246/aiti.v17i2.118-129.
- [3] V. N. Sari, L. Y. Astri, and E. Rasywir, "Analisis Dan Penerapan Algoritma Naive Bayes Untuk Evaluasi Kinerja Karyawan Pada PT. PELITA WIRA Sejahtera," *J. Ilm. Mhs. Tek. Inform.*, vol. 2, no. 1, pp. 53-68, 2020, [Online]. Available: <http://ejournal.stikom-db.ac.id/index.php/jimti/article/view/805/590>.
- [4] D. Ardiada, P. A. Ariawan, and M. Sudarma, "Evaluation of Supporting Work Quality Using K-Means Algorithm," *IJEET Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 3-6, 2018.
- [5] H. Dhika, "Kajian Perancangan Rule Kenaikan Jabatan Pada Pt. Abc," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 6, no. 2, p. 217, 2015, doi: 10.24176/simet.v6i2.455.
- [6] M. Bakri, "Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means," *J. Teknoinfo*, vol. 11, no. 1, p. 6, 2017, doi: 10.33365/jti.v11i1.3.
- [7] Aulia, D., Safii, M., Suhendro, D., Studi, M. P., Informasi, S., Tunas Bangsa, S., and Tunas Bangsa, A. Penerapan Algoritma K-Means

- dalam Proses Clustering Penilaian Kinerja ASN. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, 6, 47, 2021. <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [8] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
- [9] S. Regina, E. Sutinah, and N. Agustina, "Clustering Kualitas Kinerja Karyawan Pada Perusahaan Bahan Kimia Menggunakan Algoritma K-Means," *J. MEDIA Inform.*, vol. 5, pp. 573–582, 2021, doi: 10.30865/mib.v5i2.2909.
- [10] I. D. Nirmala and P. D. Atika, "Implementation of K-Means Algorithm As a Clustering Method for Selecting Achievement Students Based on Academic Grade," *J. Pilar Nusa Mandiri*, no. Ningrum 2009, pp. 199–204, 2020, doi: 10.33480/pilar.v16i2.1575.
- [11] D. Alfiandi, E. Ernawati, and E. P. Purwandari, "Implementasi K-Means Clustering dan Pemetaan Pemukiman Kumuh di Kota Bengkulu Berbasis Web," *Rekursif J. Informatika*, vol. 6, no. 2, 2018, [Online]. Available: <https://ejournal.unib.ac.id/index.php/rekursif/article/view/4252>.
- [12] Henderi, A. Sunarya, Zakaria, S. L. Nurmika, and N. Asmainah, "Evaluation model of students learning outcome using k-means algorithm," *J. Phys. Conf. Ser.*, vol. 1477, no. 2, 2020, doi: 10.1088/1742-6596/1477/2/022027.
- [13] S. L. B. Ginting and A. A. Permana, "Hybrid Classifier System: Support Vector Machines Dikombinasikan dengan K-Nearest Neighbors untuk Menentukan Kelayakan Nasabah Bank dalam Pengajuan Kredit," *Komputika J. Sist. Komput.*, vol. 7, no. 1, pp. 9–16, 2018, doi: 10.34010/komputika.v7i1.1402.
- [14] R. I. Fajriah, H. Sutisna, and B. K. Simpony, "Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi," *Inform. Bina Sarana Bsi, Univ.*, vol. 4, no. 1, pp. 36–49, 2019.
- [15] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019.
- [16] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *JTIK (Jurnal Tek. Informatika)*, vol. 5, no. 1, 2021.