

## Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional

*Akhmad Upi Fitriyadi<sup>1</sup>; Ana Kurniawati<sup>2</sup>*

<sup>1,2</sup>Universitas Gunadarma

<sup>1</sup>ahmaduvi99@mail.com

### **ABSTRAK**

Penilaian kinerja dilakukan untuk mengukur kinerja seorang karyawan terhadap pekerjaan yang dilakukan. Perusahaan Perumahan Nasional melakukan penilaian kinerja terhadap karyawan setiap 6 bulan sekali, yang melibatkan semua karyawan, baik karyawan tetap maupun karyawan kontrak. Tujuan dari penelitian ini adalah untuk melakukan analisis terhadap kinerja algoritma K-Means dan algoritma K-Medoids dalam melakukan proses clustering. Clustering akan dikelompokkan menjadi 4 cluster yaitu: tingkat kinerja sangat baik, tingkat kinerja baik, tingkat kinerja cukup dan tingkat kinerja kurang baik. Proses clustering akan dilakukan menggunakan software rapidminer. Pengukuran performa algoritma dalam rapidminer dilakukan menggunakan metode Confusion Matrix dengan parameter accuracy, recall dan precision. Dari penelitian yang dilakukan diperoleh hasil algoritma K-Means mempunyai tingkat accuracy 56%, presision 25% dan recall 60%, sedangkan algoritma K-Medoids mempunyai tingkat akurasi 14%, presision 25% dan recall 25%. Dengan hasil tersebut dapat disimpulkan bahwa algoritma K-Means mempunyai performa lebih baik bila dibandingkan dengan algoritma K-Medoids, karena mempunyai tingkat akurasi dan recall lebih tinggi bila dibandingkan dengan algoritma K-Medoids.

**Kata kunci:** *Data Mining, Algoritma K-Means, Algoritma K-Medoids*

### **ABSTRACT**

Performance appraisal must be carried out to measure the performance of an employee. Perumahan Nasional Company conduct employee performance appraisals every 6 months, that involves all employees, both permanent and contract employees. The purpose of this research is to analyze the performance of the K-Means algorithm and the K-Medoids algorithm in working on the clustering process. The clustering will be grouped into 4 clusters of performance level, that is: very good, good, adequate and poor. The Clustering process will be carried out using Rapidminer software. Measurement of algorithm performance in rapidminer is processed using the Confusion Matrix method with accuracy, recall and precision parameters. From the research results, the K-Means algorithm have an accuracy level of 56%, Precision 25% and Recall 60%, while the K-Medoids algorithm has an accuracy rate of 14%, Precision 25% and Recall 25%. The conclusion is that the K-Means algorithm has better performance when compared to the K-Medoids algorithm, because it has a higher level of accuracy and recall when compared to the K-Medoids algorithm.

**Keywords:** *Data Mining, K-Means Algorithm, K-Medoids Algorithm*

## 1. PENDAHULUAN

Sumber daya manusia merupakan suatu faktor yang sangat penting dalam suatu organisasi atau perusahaan. Semakin banyak karyawan yang dimiliki oleh suatu perusahaan, akan semakin berat beban Divisi Sumber Daya Manusia (SDM) dalam menjalankan tugas untuk melakukan pengelolaan karyawan. Untuk mengukur kinerja karyawan terhadap perusahaan, Divisi SDM telah melakukan penilaian kinerja terhadap masing-masing karyawan setiap 6 bulan, namun hasil dari penilaian tersebut belum dilakukan pengelompokan sesuai dengan kelas-kelas tertentu.

Untuk mempermudah Divisi SDM, dalam melakukan monitoring terhadap hasil penilaian kinerja karyawan yang telah dilakukan, diperlukan suatu metode pengelompokan data untuk menelompokkan data penilaian kinerja karyawan sesuai dengan tingkat kinerjanya. Salah satu metode pengelompokan data adalah clustering, clustering merupakan suatu metode pengelompokan data, dengan menggolongkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya. Dari beberapa algoritma yang ada dalam clustering, peneliti akan menggunakan algoritma K-Means dan K-Medoids sebagai algoritma yang akan dibandingkan dalam proses pengelompokan data kinerja karyawan.

Dalam melakukan perbandingan kinerja algoritma K-Means dan K-Medoids, peneliti akan menggunakan 3 parameter untuk melakukan pengukuran dari kinerja kedua algoritma, parameter yang digunakan adalah: Accuracy, Recall dan Precision. Dengan banyaknya parameter yang digunakan untuk melakukan pengukuran kinerja algoritma, diharapkan dapat memberikan hasil yang lebih akurat bila dibandingkan dengan hanya menggunakan 1 parameter. Dari hasil penelitian tersebut diharapkan dapat menentukan algoritma mana yang lebih baik, apabila digunakan dalam melakukan pengelompokan data kinerja karyawan.

Sebelum melakukan penelitian, peneliti melakukan review terhadap beberapa jurnal penelitian terdahulu yang dilakukan oleh peneliti-peneliti sebelumnya. Adapun jurnal-jurnal yang menjadi kajian pustaka dalam penelitian ini antara lain:

- a. Pada Penelitian Nurhayati, dkk (2018) yang berjudul “*Analysis of K-Means and K-Medoids’s Performance Using Big Data Technology*”. Penelitian ini membandingkan algoritma K-Means dan K-Medoids dengan menguji data menggunakan aplikasi berbasis Java, Hadoop, dan Hive. Parameter yang digunakan dalam pengujian ini adalah akurasi, eksekusi. Dari pengujian tersebut diperoleh kesimpulan bahwa algoritma K-Medoids lebih baik bila dibandingkan dengan algoritma K-Means.
- b. Pada Penelitian Agnis Nanda, dkk (2016) yang berjudul “*Comparative Study between Parallel K-Means and Parallel K-Medoids using Message Passing Interface (MPI)*”. Penelitian ini menganalisis hasil dari kinerja waktu algoritma K-Means dan K-Medoids menggunakan High Performance Computing (HPC) dan untuk memparalelkan algoritma K-Means dan K-Medoids menggunakan pustaka Message Passing Interface (MPI). Hasil penelitian menunjukkan bahwa algoritma K-Means memberikan Jumlah Kuadrat Kesalahan (SSE) lebih kecil daripada K-Medoids sedangkan algoritma paralel yang menggunakan MPI memberikan komputasi yang lebih cepat.
- c. Pada Penelitian Novita Lestari Anggreini, dkk (2020) yang berjudul “*Komparasi Algoritma K-Means dan K-Medoids Untuk Menangani Strategi Promosi di Politeknik TEDC Bandung*”. Penelitian ini membahas mengenai perbandingan antara Algoritma K-Means dan Algoritma K-Medoids dalam klusterisasi, dan hasilnya dapat diusulkan menjadi pertimbangan untuk memutuskan dalam menangani strategi promosi di Politeknik TEDC Bandung dalam kegiatan Penerimaan Mahasiswa Baru.

- d. Pada Penelitian Ninda Nurul Rhamadani, dkk (2020) yang berjudul “*Implementasi Algoritma K-Means dan K-Medoids Dalam Pengelompokan Nilai Ujian Nasional Tingkat SMA*”. Pada penelitian ini akan melakukan pengelompokan sekolah menjadi 3 kategori yaitu baik, sedang dan cukup untuk memudahkan peserta didik mendapatkan informasi kategori sekolah mereka dengan menerapkan algoritma K-Means dan K-Medoid. Hasil pengelompokan menggunakan algoritma k-means menghasilkan kluster baik sebanyak 14 anggota, kluster sedang 46 anggota dan kluster cukup 49 anggota. Lalu, pada algoritma kmedoids diperoleh hasil kluster baik 27 anggota, kluster sedang 43 anggota dan kluster cukup 39 anggota.
- e. Pada Penelitian Rima Dias Ramadhani, dkk (2017) yang berjudul “*Evaluasi K-Means dan K-Medoids pada Dataset Kecil*”. Pada penelitian ini akan membahas mengenai komparasi hasil evaluasi pada K-Means dan K-Medoids menggunakan dataset yang berukuran kecil yaitu dataset Iris dan Wine. Davies Bouldin Index digunakan pada penelitian ini untuk mengukur kesamaan dari ukuran kluster berdasarkan penyebaran data pada kluster dan ketidaksamaan ukuran kluster. Berdasarkan hasil eksperimen, K-Means menunjukkan hasil evaluasi yang lebih baik dibandingkan dengan K-Medoids dalam menangani dataset dengan ukuran kecil. Hal ini ditunjukkan dengan hasil evaluasi pada dataset Iris dengan menggunakan K-Means yaitu sebesar 0.662 dan pada dataset Wine menunjukkan hasil evaluasi sebesar 0.534.

Dari penelitian-penelitian yang telah dilakukan oleh peneliti sebelumnya, sebagian besar peneliti hanya menggunakan 1 parameter yang digunakan dalam melakukan pengukuran terhadap kinerja suatu algoritma. Dalam penelitian ini peneliti akan menggunakan 3 parameter untuk melakukan perbandingan terhadap performa algoritma K-Means dan K-Medoids.

Berdasarkan permasalahan yang ada dalam Divisi SDM, tujuan dari penelitian ini adalah:

- a. Melakukan analisa terhadap kinerja algoritma K-Means dan K-Medoids dalam melakukan proses clustering data.
- b. Memberikan rekomendasi kepada divisi SDM hasil dari analisa kinerja kedua algoritma tersebut.

### 1.1. Data Mining

Istilah Data Mining digunakan untuk menjelaskan atau memaparkan penemuan ilmu pengetahuan di dalam database. Data mining merupakan serangkaian proses dalam pencarian pola, hubungan, penggalian nilai tambah dari data dan informasi yang berukuran besar berupa pengetahuan dengan tujuan menemukan hubungan dan menyederhanakan data agar diperoleh informasi yang dapat dipahami dan bermanfaat dengan bantuan ilmu statistik dan matematika [1, 3, 12, 17, 19].

### 1.2. Clustering

Clustering adalah salah satu metode data mining yang bersifat tanpa arahan (unsupervised). Metode unsupervised yaitu metode yang diterapkan tanpa adanya latihan (training) dan guru (teacher) serta tidak memerlukan target output. Dalam data mining ada dua jenis metod clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering (Santosa, 2007) [5, 6, 7, 9, 18].

### 1.3. K-Means

Algoritma K-Means merupakan teknik pengelompokan data dengan cara mempartisi data ke dalam beberapa cluster dengan menetapkan sejumlah objek data terdekatnya [1, 4, 15, 16].

Berikut ini langkah-langkah clustering yang terdapat pada algoritma K-Means adalah: 1) Tentukan k sebagai jumlah cluster yang dibentuk, 2) Tentukan k centroid (titik pusat cluster) awal secara random, 3) Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster, 4) Alokasikan masing-masing objek ke dalam centroid yang paling dekat, 5) Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan, 6) Ulangi langkah 3 jika posisi centroid baru tidak sama.

### 1.4. K-Medoids

Algoritma K-Medoids atau Partitioning Around Medoids (PAM) adalah algoritma clustering yang mirip dengan K-Means. Perbedaan dari kedua algoritma ini yaitu algoritma K-Medoids atau PAM menggunakan objek sebagai perwakilan (medoid) sebagai pusat cluster untuk setiap cluster [8, 10, 11].

Langkah-langkah algoritma K-Medoids adalah: 1) Inisialisasi pusat cluster sebanyak k (jumlah cluster), 2) Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance, 3) Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru, 4) Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru, 5) Hitung total simpangan (S) dengan menghitung nilai total distance baru–total distance lama, 6) Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster beserta anggota cluster masing-masing.

### 1.5. Rapidminer

Rapidminer merupakan perangkat lunak yang bersifat terbuka (open source) yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com. Rapidminer adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapidminer memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. Rapidminer merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. Rapidminer ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi [2, 13, 14, ].

### 1.6. Confusion matrix

Confusion Matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma machine learning. Berdasarkan Confusion Matrix, bisa ditentukan parameter yang digunakan untuk mengevaluasi suatu algoritma dalam suatu machine learning (rapidminer). Dalam penelitian ini parameter yang akan digunakan untuk mengukur performance algoritma dalam machine learning (rapidminer) adalah *Accuracy*, *Precision*, *Recall* [13].

## 2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan pada penelitian ini adalah metode analisis deskriptif dengan pendekatan kuantitatif, artinya penelitian ini menekankan analisa pada data-data kinerja karyawan yang terdapat di Divisi SDM sebagai data yang akan digunakan dalam penelitian. Dalam

melakukan penelitian, akan dilakukan uji coba pemrosesan data sebanyak 1 kali dengan menggunakan data sampel 20 data kinerja karyawan yang diambil secara acak.

## 2.1. Observasi

Observasi dilakukan dengan cara mendapatkan informasi secara langsung, terkait hal-hal yang berhubungan dengan penilaian kinerja. Dalam melakukan observasi peneliti menanyakan secara langsung kepada karyawan Divisi SDM, yang bertugas melakukan penilaian kinerja. Informasi yang diperoleh peneliti selama melakukan observasi adalah: **1)** Informasi tentang jumlah karyawan, bahwa Perusahaan Perumahan Nasional mempunyai jumlah karyawan sebanyak 1520 karyawan, **2)** Pelaksanaan penilaian kinerja karyawan dilakukan per semester, semester 1 dilakukan pada bulan Januari-Juni dan semester 2 bulan Juli-Desember, **3)** Informasi tentang cara pelaksanaan penilaian kinerja, bahwa penilaian kinerja diisi oleh masing-masing karyawan, dan akan dinilai oleh atasan, **4)** Dalam pelaksanaan penilaian kinerja karyawan ada banyak parameter-parameter yang berpengaruh terhadap nilai penilaian kinerja, **5)** Untuk melakukan perhitungan penilaian kinerja, Perusahaan mempunyai formula khusus, **6)** Penilaian kinerja digunakan untuk mengukur tingkat kinerja karyawan, untuk karyawan yang mempunyai nilai kinerja rendah akan diberikan pelatihan khusus, untuk meningkatkan kinerja karyawan tersebut, **7)** Hasil dari penilaian kinerja belum dilakukan penelompokan secara khusus.

## 2.2. Pengumpulan Data

Penumpulan data dilakukan untuk mendapatkan data-data yang digunakan dalam penelitian. Dalam melakukan pengumpulan data, peneliti meminta data secara langsung kepada karyawan Divisi SDM yang bertugas melakukan penilaian kinerja. Data yang diperoleh selama melakukan pengumpulan data adalah berupa data excel, yang berisi data penilaian kinerja secara detail.

Gambar 1 merupakan contoh data penilaian kinerja yang diperoleh penulis selama melakukan pengumpulan data. Dalam data tersebut menunjukkan detail data penilaian kinerja dalam setiap parameter yang digunakan dalam melakukan pengukuran kinerja karyawan.

Nama Suradi  
Departemen Subag SDM, Dana & Umum  
Jabatan Kasir Regional

A	KPI (KINERJA)	BOBOT	TARGET	REALISASI	NILAI	BOBOT x NILAI
<b>Dari Strategi</b>						
1	Pendapatan Usaha (Rp. Miliar)	12.5%	204,405.50	9,918.23	5%	1%
2	Net Profit Margin (Rp. Miliar)	12.5%	61,221	-12,135.70	-20%	0%
3	Cash In Operasional to Sales (%)	10.0%	50%	44%	88%	9%
4	COGS/Sales (%) Regional	5.0%	72%	72%	100%	5%
5	Cash to Cash Cycle Time (hari)	10.0%	599	599	100%	10%
6	Saldo piutang terhadap realisasi pencairan	10.0%	60%	17%	29%	3%
<b>Dari Job Desc. (40%)</b>						
1	% Ketepatan transfer dana operasional	20.0%	100%	95%	95%	19%
2	% ketepatan pelaporan Mingguan	10.0%	100%	95%	95%	10%
3	% ketepatan penggunaan kas kecil	10%	90%	85%	94%	9%
100.00%						65.23%
B	KOMPETENSI	BOBOT	TARGET	REALISASI	NILAI	BOBOT x NILAI
1	Pekerjaan	20%	100	90	90%	18%
2	Kehadiran	10%	100	90	90%	9%
3	Penampilan / Pelayanan	10%	100	87	87%	9%
4	Agresifitas / Hasrat Bersaing	20%	100	85	85%	17%
5	Problem Solving	20%	100	87	87%	17%
6	Orientasi Kualitas Hasil Kerja	20%	100	90	90%	18%
100%						88.10%

**Gambar 1.** Data Penilaian Kinerja Karyawan

### 2.3. Transformasi Data

Transformasi data dilakukan untuk menyesuaikan format data yang dapat diproses oleh software Rapidminer, karena rapidmier mempunyai formas kusus agar data dapat di proses.

Tabel 1 Menunjukkan format data excel yang digunakan peneliti untuk memasukan data kinerja karyawan kedalam software rapidminer. Dalam tabel tersebut menunjukkan data kinerja karyawan yang telah dilakukan transformasi data dan siap untuk dilakukan pemrosesan data.

**Tabel 1.** Data Penilaian Kinerja Hasil Transformasi

No	Kinerja		Kompetensi					
	Strategi	Job Desc	Pekerjaan	Kehadiran	Penampilan	Aggresifitas	Problem Solving	Hasil Kerja
k1	29	39	19	10	10	19	18	18
k2	27	40	18	9	9	17	18	18
k3	29	39	18	9	9	18	18	18
k4	29	38	20	10	9	20	19	19
k5	23	40	18	9	9	19	19	19
k6	23	36	18	10	9	18	18	18
k7	27	40	19	9	9	18	18	19
k8	27	40	19	9	9	18	18	19
k9	27	37	19	10	10	19	19	19
k10	27	40	19	9	9	18	18	19
k11	37	35	18	9	9	17	17	17
k12	26	37	19	9	9	19	18	19
k13	26	36	18	9	9	18	17	18
k14	26	34	17	9	9	17	16	17
k15	29	34	18	9	9	18	17	18
k16	26	34	18	9	9	17	17	18
k17	26	36	18	9	9	18	17	18
k18	26	36	17	9	9	16	17	17
k19	31	35	18	9	9	17	17	18
k20	22	39	19	10	10	19	19	19

Data yang akan digunakan dalam penelitian ini adalah data penilaian kinerja karyawan semester 2 tahun 2019. Sedangkan data sampel yang aan digunakann berjumlah 20 data sampel yang telah diambil secara acak. Dalam proses penilaian kinerja ada beberapa parameter yang akan digunakan selama proses penilaian kinrja, parameter tersebut adalah strategi, job desc, pekerjaan, kehadiran, penampilan, aggresifitas, problem solving dan hasil kerja.

## 3. PENGUJIAN DAN HASIL

Penelitian ini akan melakukan perbandingan terhadap kinerja algoritma K-Means dan algoritma K-Medoids, dimana algoritma yang lebih baik akan direkomendasikan peneliti kepada perusahaan untuk dijadikan sebagai rujukan dalam melakukan penelompokan data.

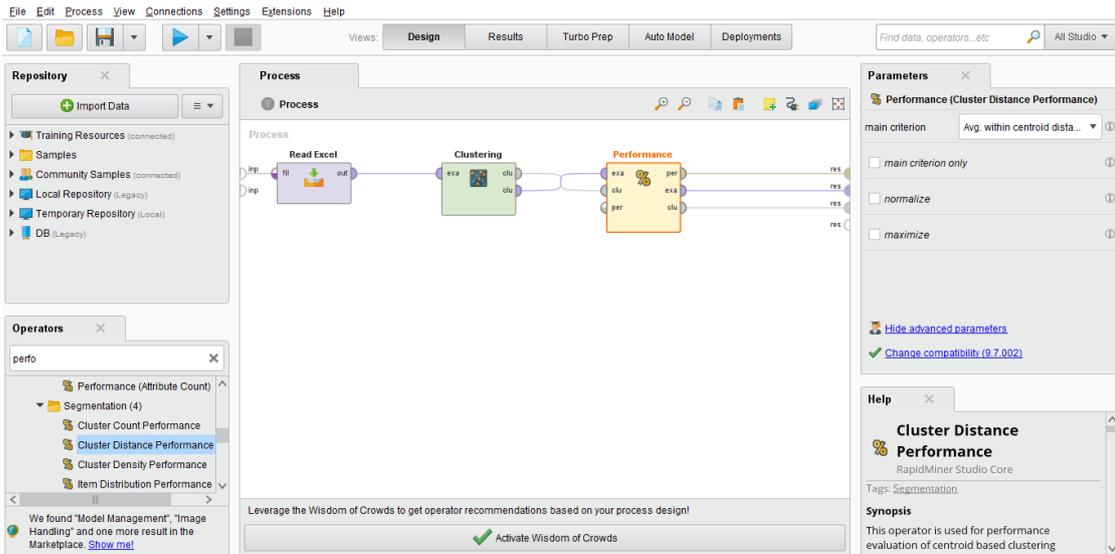
### a. Pengujian Menggunakan Rapidminer

Dalam melakukan pengujian menunakan rapid miner, maka peneliti akan memasukan data dalam format excel, untuk selanjutnya dilakukan pemrosesan data. Penujian ini dilakukan dua kali yaitu denan algoritma K-Means dan algoritma K-Medoids.

Tahapan proses pengolahan data menunakan rapidminer adalah:

#### 1. Desain Input Proses

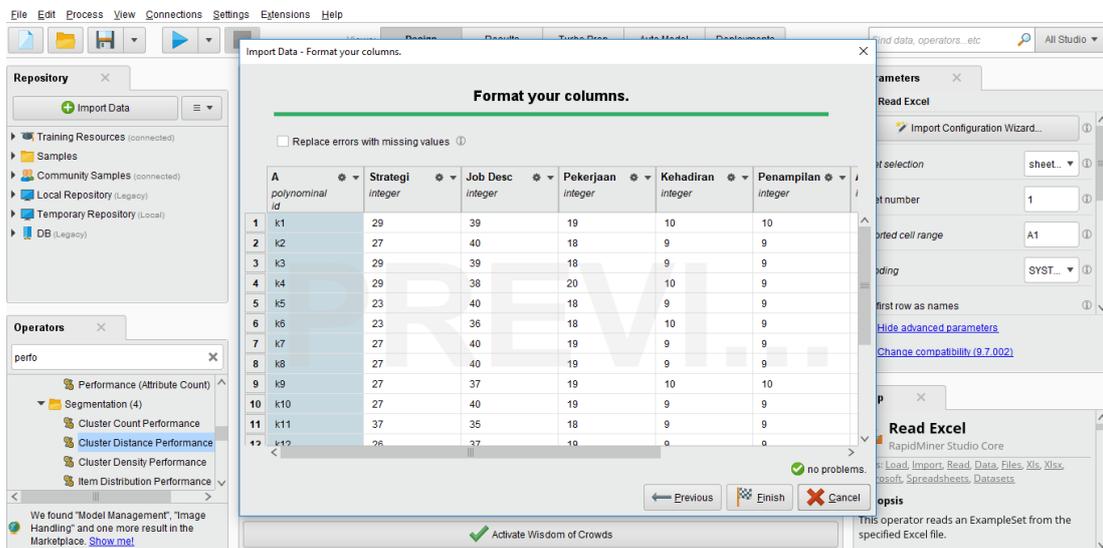
Sebelum melakukan pengolahan data menggunakan rapidminer tahap pertama yang dilakukan adalah membuat desain input proses. Desain input proses ini terdiri dari 3 operator yaitu, operator Read excel, algoritma clustering dan performance. Gambar 2 menunjukkan desain proses yang akan dijalankan dalam rapidminer.



Gambar 2. Desain Input Proses

2. Proses Pengambilan Data

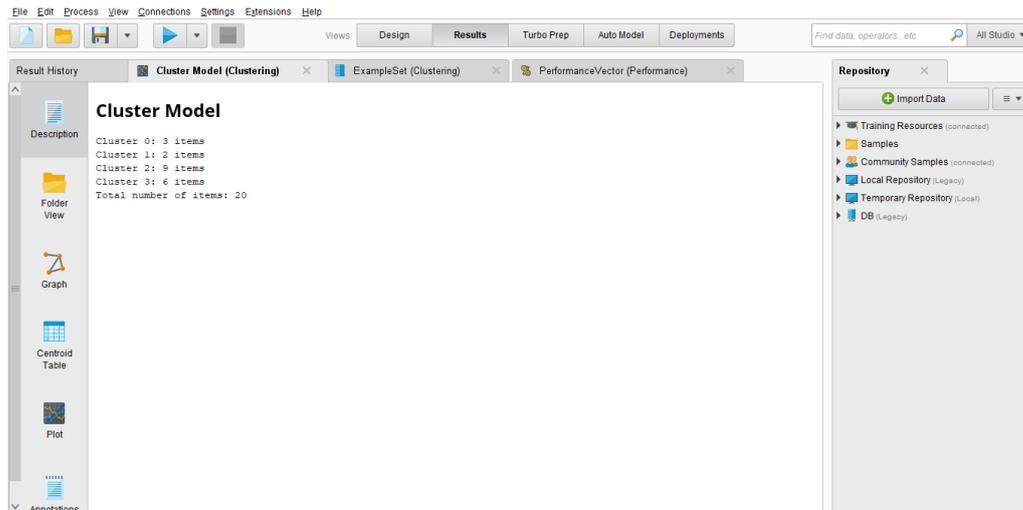
Proses pengambilann data merupakan tahapan yang dilakukan untuk memasukan data excel yang telah disiapkan, kedalam software rapidminer. Pada tahap ini data akan dimasukan ke dalam rapidminer, untuk selanjutnya dilakukan pemrosesan. Gambar 3 merupakan proses import data excel ke dalam *software* rapidminer.



Gambar 3. Proses Import Data Excel

3. Hasil Clustering

Output pertama yang dihasilkan setelah tahap proses menggunakan rapidminer adalah output tentang jumlah data dalam stiap cluster. Output ini memberikan informasi secara global tentang jumlah data dalam setiap cluster. Gambar 4 merupakan output data clustering per cluster secara global. Pada Gambar 4 kita bisa memperoleh informasi tentang jumlah data kinerja karyawan per cluster.



**Gambar 4.** Hasil Clustering

**4. Hasil Tabel Centroid**

Output ini memberikan informasi tentang titik centroid dalam setiap cluster, dari hasil proses data yang telah dilakukan. Gambar 5 menunjukkan data centroid baru yang terbentuk setelah proses clustering data menggunakan rapidminer.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Strategi	22.667	34	27.556	26.500
Job Desc	38.333	35	38.889	35
Pekerjaan	18.333	18	18.889	17.667
Kehadiran	9.667	9	9.333	9
Penampilan	9.333	9	9.222	9
Agresifitas	18.667	17	18.444	17.333
Problem Solving	18.667	17	18.222	16.833
Hasil Kerja	18.667	17.500	18.667	17.667

**Gambar 5.** Hasil Tabel Centroid

**b. Hasil Pengujian**

Dari hasil pengujian kedua algoritma yang telah dilakukan diperoleh hasil pengelompokan data kinerja karyawan dalam setiap cluster. Tabel 2 memberikan informasi mengenai jumlah data dalam setiap cluster dari kedua algoritma. Dari hasil pengujian kedua algoritma menunjukkan, meskipun data kinerja karyawan yang diproses sama, ternyata setelah dilakukan pengelompokan menggunakan algoritma berbeda, ternyata memberikan hasil cluster yang berbeda pula.

**Tabel 2.** Tabel Hasil Pengujian

No	Algoritma	Cluster 0	Cluster 1	Cluster 2	Cluster 3
1	Algoritma K-Means	3	2	9	6
2	Algoritma K-Medoids	9	11	0	0

### c. Perbandingan Kinerja Algoritma

Dalam melakukan evaluasi terhadap performance algoritma machine learning (khususnya rapidminer), kita menggunakan acuan Confusion Matrix. Confusion Matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma machine learning. Berdasarkan Confusion Matrix, bisa ditentukan parameter yang *digunakan* untuk mengevaluasi suatu algoritma dalam suatu machine learning (rapidminer). Dalam penelitian ini parameter yang akan digunakan untuk mengukur performance algoritma dalam machine learning (rapidminer) adalah Accuracy, Precision, Recall.

Dari hasil pengujian yang telah dilakukan, dapat dilakukan perhitungan Confusion Matrix dari kedua algoritma tersebut adalah:

#### 1. Algoritma K-Means

Tabel 3 merupakan tabel matrix Confusion Matrix dengan menggunakan algoritma K-Means.

**Tabel 3.** Tabel Confusion Matrix K-Means

Analisa Dengan		K-Means	
		Cluster 0	Cluster 1
K-Medoids	Cluster 0	3	9
	Cluster 1	2	11

#### 2. Algoritma K-Medoids

Tabel 3 merupakan tabel matrix Confusion Matrix dengan menggunakan algoritma K-Medoids.

**Tabel 4.** Tabel Confusion Matrix K-Medoids

Analisa Dengan		K-Means	
		Cluster 0	Cluster 3
K-Medoids	Cluster 0	3	9
	Cluster 3	9	0

### i. Accuracy

Untuk mengetahui tingkat akurasi suatu algoritma dalam melakukan pemrosesan data, maka harus dilakukan perhitungan dengan persamaan sebagai berikut:

$$\text{Accuracy} = ((TP+TN)/(TP+FP+FN+TN)) \times 100\%$$

(1)

Dari pengujian kinerja karyawan yang telah dilakukan, maka peneliti akan melakukan perhitungan akurasi dari hasil pemrosesan data tersebut. Berikut merupakan detail dari perhitungan accuracy dengan menggunakan confusion matrix:

#### 1. Algoritma K-Means

$$\begin{aligned} \text{Accuracy} &= ((3+11)/(3+9+2+11)) \times 100\% \\ &= (14/25) \times 100\% \\ &= 0.56 \times 100\% \\ &= 56\% \end{aligned}$$

#### 2. Algoritma K-Medoids

$$\begin{aligned} \text{Accuracy} &= ((3+0)/(3+9+9+0)) \times 100\% \\ &= (3/21) \times 100\% \\ &= 0.14 \times 100\% \\ &= 14\% \end{aligned}$$

Dengan menggunakan rumus confusion matrix diperoleh, nilai accuracy dari kedua algoritma. Algoritma K-Means mempunyai nilai accuracy 56% sedangkan algoritma K-Medoids mempunyai nilai accuracy 14%.

## ii. Precision

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Berikut merupakan detail perhitungan precision dari kedua algoritma.

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

(2)

### 1. Algoritma K-Means

$$\begin{aligned}\text{Precision} &= (3/(3+9)) \times 100\% \\ &= (3/12) \times 100\% \\ &= 0.25 \times 100\% \\ &= 25\%\end{aligned}$$

### 2. Algoritma K-Medoids

$$\begin{aligned}\text{Precision} &= (3/(3+9)) \times 100\% \\ &= (3/12) \times 100\% \\ &= 0.25 \times 100\% \\ &= 25\%\end{aligned}$$

Dengan menggunakan rumus confusion matrix diperoleh, nilai precision dari kedua algoritma. Algoritma K-Means mempunyai nilai precision 25% sedangkan algoritma K-Medoids mempunyai nilai accuracy 25%.

## iii. Recall

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Berikut merupakan hasil perhitungan dengan menggunakan recall.

$$\text{Recall} = (\text{TP}) / (\text{TP} + \text{FN})$$

(3)

### 1. Algoritma K-Means

$$\begin{aligned}\text{Recall} &= (3/(3+2)) \times 100\% \\ &= (3/5) \times 100\% \\ &= 0.6 \times 100\% \\ &= 60\%\end{aligned}$$

### 2. Alortima K-Medoids

$$\begin{aligned}\text{Recall} &= (3/(3+9)) \times 100\% \\ &= (3/12) \times 100\% \\ &= 0.25 \times 100\% \\ &= 25\%\end{aligned}$$

Dengan menggunakan rumus confusion matrix diperoleh, nilai recall dari kedua algoritma. Algoritma K-Means mempunyai nilai recall 60% sedangkan algoritma K-Medoids mempunyai nilai accuracy 25%.

## 4. KESIMPULAN DAN SARAN

### a. Kesimpulan

Dari hasil perhitungan yang telah dilakukan dengan menggunakan metode Confusion Matrix, diperoleh hasil bahwa: Dengan menggunakan parameter accuracy algoritma K-Means mempunyai

nilai 56% sedangkan K-Medoids mempunyai nilai 14%, Dengan menggunakan parameter recall algoritma K-Means mempunyai nilai 60% sedangkan K-Medoids mempunyai nilai 25%, Dengan menggunakan parameter precision algoritma K-Means mempunyai nilai 25% sedangkan K-Medoids mempunyai nilai 25%. Hasil dari perhitungan yang dilakukan dapat ditarik kesimpulan bahwa algoritma K-Means lebih baik bila dibandingkan algoritma K-Medoids, karena mempunyai tingkat akurasi dan recall lebih tinggi bila dibandingkan dengan algoritma K-Medoids. Dari hasil tersebut peneliti akan merekomendasikan algoritma K-Means sebagai algoritma yang akan digunakan dalam melakukan clustering data kinerja karyawan kepada Divisi SDM.

#### b. Saran

Penelitian ini baru melakukan clustering terhadap karyawan di level staf, sementara di perusahaan Perumahan Nasional masih banyak lagi karyawan dengan jabatan-jabatan yang lain. Untuk dapat melihat clustering data secara global perlu dilakukan penelitian mengenai clustering data kinerja karyawan secara keseluruhan.

Algoritma yang digunakan untuk melakukan clustering data sangat banyak, oleh karena itu perlu dilakukan perbandingan clustering karyawan dengan algoritma yang lain, untuk mengetahui performance dari algoritma tersebut. Misalnya algoritma X-Means, Fuzzy C-Means dan Mixture Modelling. Selain menggunakan algoritma yang berbeda, penelitian juga bisa dilakukan menggunakan menggunakan machine learning yang lain misalnya WEKA, Rattle, Orange dan KNIME.

#### DAFTAR PUSTAKA

- [1] I. H. Rifa, H. Pratiwi, and R. Respatiwan, "Clustering of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms," *Media Stat.*, vol. 13, no. 2, pp. 194–205, 2020, doi: 10.14710/medstat.13.2.194-205.
- [2] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional," *J. Ilm. FIFO*, vol. 11, no. 2, p. 208, 2019, doi: 10.22441/fifo.2019.v11i2.010.
- [3] Y. H. Susanti and E. Widodo, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," vol. 1, no. 1, pp. 116–122, 2017.
- [4] Mediana, E. D. Madyatmadja, and E. Miranda, "Application of K-Means and K-Medoids Clustering Pada Data Internet Banking Di Bank Xyz," pp. 349–356, 2018.
- [5] P. Kumar and D. Sirohi, "Comparative analysis of FCM and HCM algorithm on Iris data set," *Int. J. Comput. Appl.*, vol. 5, no. 2, pp. 33–37, 2010, doi: 10.5120/888-1261.
- [6] C. K. Raj, "Comparison of k-means, k-medoids, DBSCAN algorithms using DNA microarray dataset," *Int. J. Comput. Appl. Math.*, vol. 12, no. 1, pp. 344–355, 2017.
- [7] W. AS, M. K. Aidid, and M. Nusrang, "Pengelompokan Kabupaten/Kota Provinsi Sulawesi Selatan dan Barat Berdasarkan Angka Partisipasi Pendidikan SMA/SMK/MA Menggunakan K-Medoid dan CLARA," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 1, no. 3, p. 48, 2019, doi: 10.35580/variansiunm12899.
- [8] R. Goejantoro, "Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018) Comparison," vol. 10, no. April 2018, pp. 143–152, 2019.
- [9] S. Nirmal, "Comparative study between k-means and k-medoids clustering algorithms," *Int. Res. J. Eng. Technol.*, vol. 839, pp. 839–844, 2019, [Online]. Available: <https://www.irjet.net/archives/V6/i3/IRJET-V6I3154.pdf>.
- [10] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids

- dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [11] F. Nhita, “Comparative Study between Parallel K-Means and Parallel K-Medoids with Message Passing Interface (MPI),” *Int. J. Inf. Commun. Technol.*, vol. 2, no. 2, p. 27, 2017, doi: 10.21108/ijoict.2016.22.86.
- [12] P. Surya and I. Laurence Aroquiaraj, “Performance Analysis of K-Means and K-Medoid Clustering Algorithms Using Agriculture Dataset,” *J. Emerg. Technol. Innov. Res.*, vol. 6, no. 1, pp. 539–545, 2019.
- [13] N. Lestari Anggreini and S. Tresnawati, “Komparasi Algoritma K-Means dan K-Medoids Untuk Menangani Strategi Promosi di Politeknik TEDC Bandung,” *TEDC*, vol. 14, no. 2, pp. 120–126, 2020.
- [14] I. Algoritma and K. D. A. N. K. Dalam, “Implementasi algoritma k-means dan k-medoids dalam pengelompokan nilai ujian nasional tingkat smk,” no. Ciastech, pp. 717–726, 2020.
- [15] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, “K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir,” *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.
- [16] M. Fuady and J. Nugraha, “IMPLEMENTASI METODE K-MEANS DAN K-MEDOIDS UNTUK Jurusan Statistika , FMIPA , Universitas Islam Indonesia Isnafuady1@gmail.com ABSTRAK PENDAHULUAN Inflasi merupakan permasalahan yang selalu dan pasti dialami oleh setiap negara . Disebabkan oleh pengaruh in,” pp. 327–337, 2017.
- [17] Nurhayati, N. S. Sinatrya, L. K. Wardhani, and Busman, “Analysis of K-Means and K-Medoids’s Performance Using Big Data Technology,” *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–5, 2019, doi: 10.1109/CITSM.2018.8674251.
- [18] I. Kamila, U. Khairunnisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [19] A. Dharmarajan and T. Velmurugan, “Efficiency of k-Means and k-Medoids Clustering Algorithms using Lung Cancer Dataset,” *Int. J. Data Min. Tech. Appl.*, vol. 5, no. 2, pp. 150–156, 2016, doi: 10.20894/ijdmata.102.005.002.011.
- [20] R. D. Ramadhani and D. J. Ak, “Evaluasi K-Means dan K-Medoids pada Dataset Kecil,” *Semin. Nas. Inform. dan Apl.*, no. September, pp. 20–24, 2017.