

## REKOGNISI KARYAWAN BERDASARKAN KINERJA MENGUNAKAN ALGORITMA K-MEANS PADA APLIKASI RAPIDMINER

<sup>1</sup>Isminarti, <sup>1</sup>Yoan Elviralita, <sup>2</sup>Nanang Roni Wibowo, <sup>2</sup>Deni Kurnia, <sup>2</sup>Janizal

<sup>1</sup>Teknik Mekatronika, Politeknik Bosowa, Makassar

<sup>2</sup>Teknologi Rekayasa Mekatronika, Politeknik Enjinering Indorama, Purwakarta

<sup>1</sup>e-mail: isminarti@politeknikbosowa.ac.id

### Abstrak

Absensi karyawan menjadi parameter penting dalam suatu rekognisi di suatu perusahaan. Dataset mengenai absensi karyawan pada penelitian ini di ambil dari UCI Machine Learning Repository dengan data latih sebanyak 740 karyawan dengan beberapa kriteria penilaian kinerja dan memilih 5 atribut centroid yaitu biaya transportasi, jarak tempat tinggal ke kantor, usia, ketidakdisiplinan dan pendidikan. Penelitian ini menggunakan metode klasterisasi dengan algoritma K-Means dimana aplikasi atau perangkat lunak yang digunakan adalah RapidMiner yang berfungsi sebagai alat pembelajaran pada ilmu data mining. Hasil penelitian ini dibagi menjadi 2 skenario yaitu pertama, cluster dibagi menjadi 2 yaitu 0 dengan kinerja baik dan 1 dengan kinerja buruk. Kedua, cluster dibagi menjadi 4 dimana 0 dengan kinerja sangat baik, 1 kinerja baik, 2 kinerja cukup, 3 kinerja kurang. Dari kedua skenario tersebut dapat disimpulkan bahwa K-Means clustering mengelompokkan data yang ada ke dalam beberapa kelompok dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan menggunakan metode ini perusahaan akan mudah menentukan penghargaan/rekognisi kepada karyawannya.

**Kata Kunci:** rekognisi, data mining, dataset, K-Means, cluster

### Abstract

Employee attendance becomes an important parameter in recognition in a company. The dataset regarding employee absenteeism in this study was taken from the UCI Machine Learning Repository with data on as many as 740 employees with several performance appraisal criteria and choosing 5 attributes, namely transportation costs, distance from residence to office, age, indiscipline, and education. This study uses a clustering method with the K-Means algorithm where the application or software used is RapidMiner which functions as a learning tool in data mining science. The results of this study are divided into 2 scenarios, namely first, cluster into 2, namely 0 with good performance and 1 with poor performance. Second, the cluster is divided into 4 were 0 with good performance, 1 with good performance, 2 with moderate performance, and 3 with poor performance. From the two scenarios, it can be said that K-Means clustering groups the existing data into several groups where the data in one group have the same characteristics and has different characteristics from the data in other groups. By using this method, the company will easily determine the award/recognition for its employees.

**Key Word :** recognition, data mining, dataset, K-Means, cluster

## 1. PENDAHULUAN

Rekognisi diberikan oleh perusahaan ketika karyawan memenuhi persyaratan penilaian yang telah ditetapkan berdasarkan PA (Performance Appraisal). Penilaian PA akan mudah di ukur dengan system sehingga jika suatu perusahaan menerapkan suatu system sebagai indikator penilaian maka dengan menggunakan RapidMiner hasil yang diinginkan akan mudah didapatkan. Dalam penelitian ini penulis menggunakan dataset dari UCI machine learning dimana data yang ada ribuan namun penulis hanya mengambil 740 dataset saja[1]. Dari data yang di ambil ada beberapa nilai yang memiliki perbedaan yang signifikan oleh karena itu dibutuhkan proses normalisasi. Dengan menggunakan metode clustering menggunakan algoritma *K-Means* penelitian ini dibuat menjadi 2 skenario sehingga data ini bisa dijadikan acuan untuk pemberian rekognisi. Skenario 1 membagi kelas karyawan menjadi 2 bagian yaitu karyawan yang berhak menerima bonus dan tidak berhak menerima bonus. Skenario 2 kategori persentase pemberian bonus secara proporsional dimana kelas karyawan dibagi menjadi 4 bagian yaitu 0 untuk kinerja sangat baik, 1 kinerja baik, 2 kinerja cukup dan 3 untuk kinerja kurang. Dari kedua skenario tersebut penelitian ini bertujuan mengklusterkan karyawan untuk memberikan rekognisi baik berupa bonus maupun kenaikan gaji secara berkala berdasarkan kinerja yang terukur. Harapan dari penelitian ini adalah perusahaan secara adil berdasarkan fakta terukur dapat memberikan rekognisi kepada karyawan sehingga manfaat penelitian ini bagi karyawan adalah mampu memotivasi, menstimulasi dan membangun loyalitas karyawan, sedangkan dari sisi perusahaan memberikan kemudahan mengukur kinerja karyawan sehingga proses pengambilan keputusan lebih cepat dalam memberikan rekognisi berupa reward dan juga punishment bagi yang tidak berkompeten.

## 2. LANDASAN TEORI

### 2.1. Algoritma klasifikasi *K-Means*

*K-Means* merupakan algoritma clustering yang dimulai dengan pemilihan secara acak K, dimana K merupakan banyaknya Cluster dan K ditetapkan secara random, untuk menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Perhitungan jarak setiap data yang ada pada masing-masing centroid menggunakan rumus Euclidian sehingga ditemukan jarak yang paling dekat dari setiap data dengan centroid [2].

### 2.2. Data Mining

Biasa disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [3]. Data Mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar [4].

Algoritma *K-Means* merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering. Pengelompokan data dengan metode KMeans dilakukan dengan cara :

1. Menentukan jumlah kelompok
2. Mengalokasikan data ke dalam kelompok secara acak
3. Menghitung pusat kelompok (centroid/rata-rata) dari data yang ada pada masing-masing kelompok. Centroid setiap kelompok diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya. Jika M adalah jumlah data dalam sebuah kelompok, i menyatakan fitur ke-i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung centroid fitur ke-l adalah dinyatakan dalam persamaan 1:

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

Persamaan 1 dilakukan sebanyak p dimensi dari i=1 sampai dengan i=p.

4. Mengalokasikan masing-masing data ke centroid/rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean. Pengukuran jarak pada ruang jarak (distance space) Euclidean dapat dicari menggunakan persamaan 2 berikut ini :

$$d = \sqrt{(x_1 - x_2)^2 + (x_1 - x_2)^2} \quad (2)$$

Data dialokasikan ulang secara tegas ke kelompok yang mempunyai cetroid dengan jarak terdekat dari data tersebut [5].

K – means adalah salah satu metode data mining unsupervised. Ada dua jenis data clustering yang sering digunakan dalam proses pengelompokan data yaitu hirarki data clustering dan non hirarki data clustering. *K-Means* merupakan salah satu metode data clustering non hirarki yang mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok [6]. Adapun tujuan dari data clustering ini adalah untuk meminimalisasi objective function yang ditentukan pada saat proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster [7].

Adapun prosedur pengambilan sampel adalah sebagai berikut [8] :

1. Pertama, tetapkan nilai untuk semua data. Pada saat yang sama, pemrosesan terpadu dilakukan dalam format kata kunci.
2. Urutkan data di atas dari yang terbesar ke yang terkecil.
3. Pilih data terkecil setelah disortir sebagai titik pusat dari cluster kelas inisialisasi.

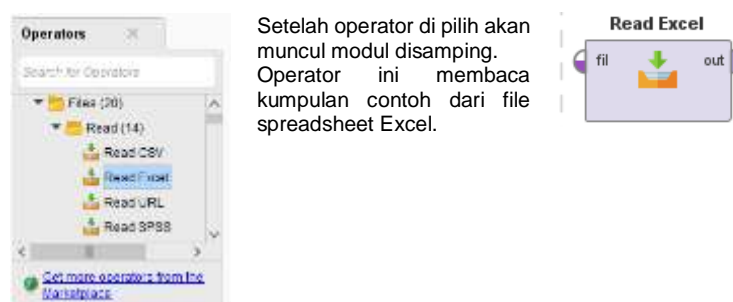
### 2.3. Clustering

Clustering atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. Cluster adalah sekelompok atau sekumpulan objek-objek data yang mirip satu sama lain dalam cluster yang sama dan tidak mirip [9].

## 3. METODE PENELITIAN

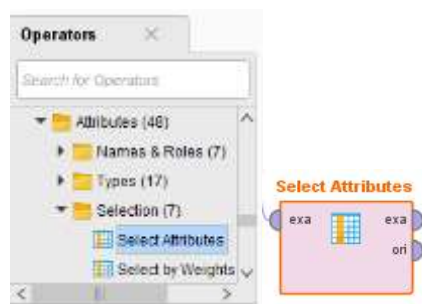
*K-Means* adalah salah satu metode pengelompokan data non hierarki yang mempartisi data ke dalam dua atau lebih bentuk kelompok. Adapun tujuan pengelompokan data ini adalah meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok. Dari beberapa definisi di atas dapat disimpulkan bahwa *K-Means* Clustering adalah metode pengelompokan yang mempartisi data ke dalam kelompok yang memiliki karakteristik sama agar tidak tumpang tindih [10].

Adapun tahapan clustering atau pengelompokan dengan metode *K-Means* pada Rapidminer adalah sebagai berikut : langkah pertama yang dilakukan adalah memilih operator pembacaan (*read*) sebagaimana ditunjukkan pada Gambar 1.



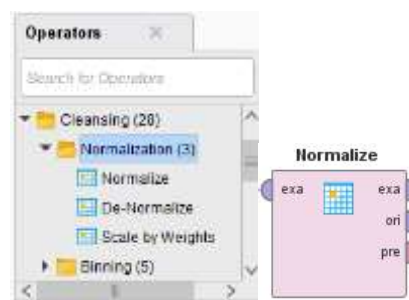
**Gambar 1.** Operator Pembacaan Dataset.

Berdasarkan Gambar 1 dapat diketahui pada tab Operator Rapidminer mendukung beberapa sumber dataset, pada penelitian ini Dataset yang digunakan adalah dataset yang berasal dari UCI Machine Learning Repository dengan judul absenteeism at work data set yang berupa file spreadsheet Excell. Langkah selanjutnya adalah menentukan attribute yang akan digunakan sebagaimana ditunjukkan pada Gambar 2.



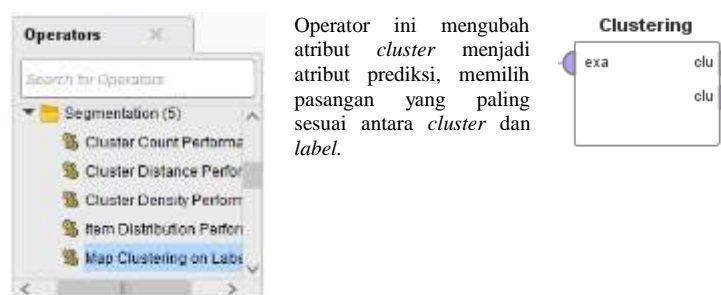
Gambar 2. Operator Seleksi Atribut.

Gambar 2 menunjukkan bahwa operator ini mengizinkan untuk memilih atribut mana yang harus menjadi bagian dari ExampleSet yang dihasilkan. Pemilihan dapat dilakukan dengan menggunakan beberapa kondisi. Setelah menentukan attribute selanjutnya adalah melakukan normalisasi sebagaimana Gambar 3.



Gambar 3. Operator Normalisasi data.

Berdasarkan Gambar 3 operator normalize dilakukan jika ada beberapa nilai atribut yang tidak terbaca dan digunakan untuk menormalkan nilai atribut untuk rentang tertentu. Langkah selanjutnya adalah melakukan modelling sebagaimana ditunjukkan pada Gambar 4.



Operator ini mengubah atribut *cluster* menjadi atribut prediksi, memilih pasangan yang paling sesuai antara *cluster* dan *label*.

Gambar 4. Operator clustering.

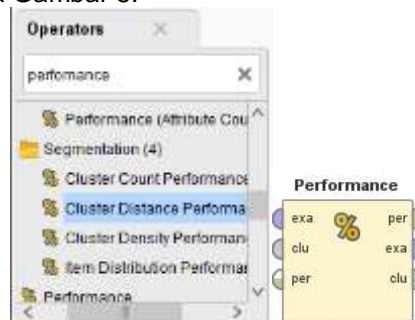
Proses Modelling pada Gambar 4 dilakukan menggunakan segmentasi metode clustering dengan algoritma *K-Means*. *K-Means Clustering* adalah salah satu “*unsupervised machine learning algorithms*” yang paling sederhana dan populer. *K-Means Clustering* adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi.

Metode *K-Means Clustering* bertujuan untuk meminimalisasikan *objective function* yang diset dalam proses *clustering* dengan cara meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya juga bertujuan untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variabel *K*. Variabel *K* sendiri adalah jumlah cluster yang diinginkan. Algoritma ini menerima masukan berupa data tanpa label kelas.

Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan *k* buah kelompok (*cluster*)

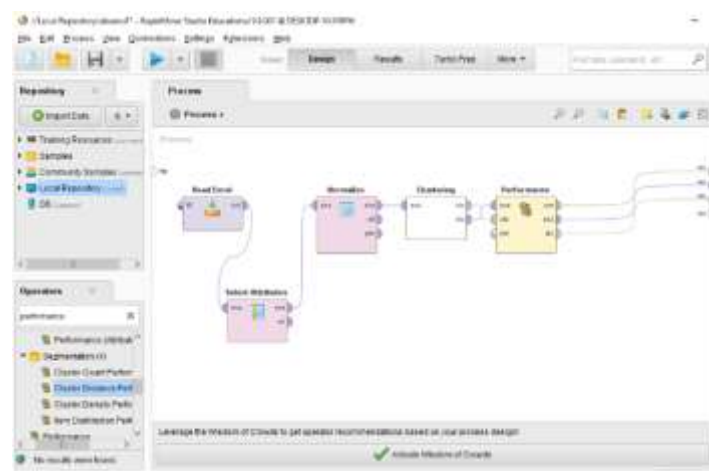
yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut.

Terdapat dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *Hierarchical* dan *Non-Hierarchical*, dan *K-Means* merupakan salah satu metode data *clustering non-hierarchical* atau *Partitional Clustering*. Setelah diperoleh data hasil *clustering* selanjutnya adalah melakukan validasi sebagaimana Gambar 5.



Gambar 5. Operator cluster distance performance.

Pada Gambar 5 menunjukan proses segmentasi untuk melakukan validasi dengan menggunakan operator *cluster distance performance* yang digunakan untuk memberikan kinerja berdasarkan *centroid cluster*. Langkah selanjutnya adalah merangkai semua modul operator yang dipilih di *tab process* sebagaimana ditunjukan pada Gambar 6.



Gambar 6. Proses Design Clustering dengan K-Means.

Berdasarkan Gambar 6 diperoleh hasil keseluruhan modul operator yang akan digunakan untuk melakukan *modelling* dengan metode *clustering* menggunakan algoritma *K-Means* pada Rapidminer dan selanjutnya diperoleh data hasil *clustering*.

#### 4. HASIL DAN PEMBAHASAN

Pada Penelitian ini terdapat 2 skenario yang akan dilakukan uji coba dengan menggunakan metode *K-Means Clustering* yaitu :

1. *Cluster* dibagi menjadi 2 yaitu :
  - a. 0 untuk kinerja baik dan
  - b. 1 untuk kinerja buruk.
2. *Cluster* dibagi menjadi 4 yaitu :
  - a. 0 untuk kinerja sangat baik,
  - b. 1 untuk kinerja baik,
  - c. 2 untuk kinerja cukup,
  - d. 3 untuk kinerja kurang

Setelah proses validasi pada Gambar 5 dilakukan kemudian di-run dan hasilnya sebagaimana ditunjukkan pada Gambar 7.



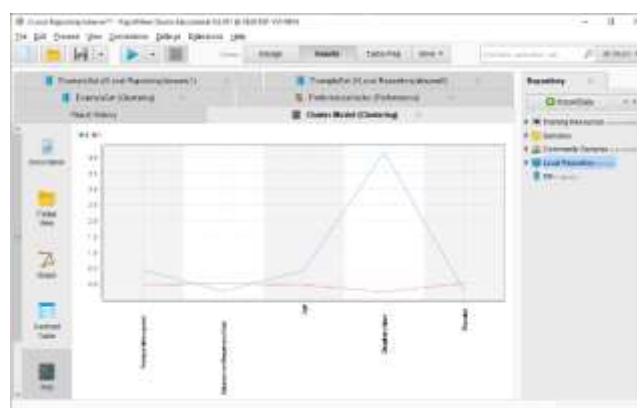
**Gambar 7.** Hasil Running Model Cluster.

Berdasarkan Gambar 7 diperoleh hasil bahwa ada 2 cluster yang digunakan dari total data 740 data dimana cluster 0 ada 40 data dan cluster 1 ada 700 data, proses ini merupakan proses untuk skenario 1 yang diujicobakan. Dengan data hasil dalam bentuk tabel centroid sebagaimana ditunjukkan Gambar 8.

Cluster	Transportation expense	Distance from residence to work	Age	Disciplinary failure	Education
Cluster 0	3.477	4.028	0.016	4.028	0.016
Cluster 1	3.477	4.028	0.016	4.028	0.016

**Gambar 8.** Tabel centroid.

Berdasarkan Gambar 8 menunjukkan dari 2 cluster dipilih 5 atribut sebagai centroid (posisi rata-rata dari semua titik dalam bentuk dua dimensi) yaitu : Transportation expense (biaya transportasi), Distance from residence to work (jarak dari tempat tinggal ke kantor), Age (umur), Disciplinary failure (ketidaksiplinan), Education (pendidikan). Untuk mendapatkan hasil dalam bentuk grafik dapat digunakan *plot system* sebagaimana ditunjukkan Gambar 9.



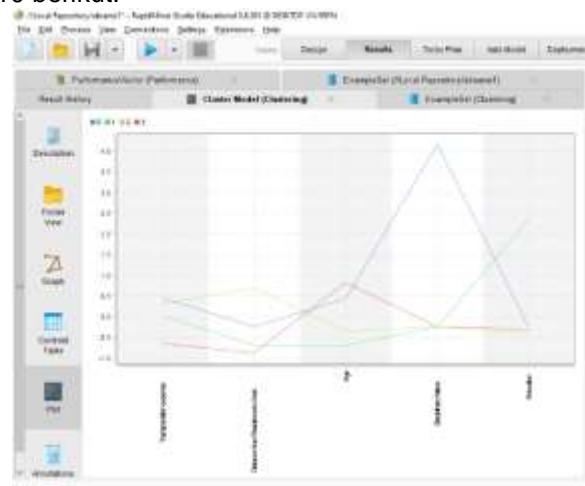
**Gambar 9.** Grafik hasil klasterisasi.

**Tabel 1.** Analisis kinerja berdasarkan cluster dan atribut.

Atribut	Cluster	
	0 (40 karyawan)	1 (700 karyawan)
Biaya Transportasi	Tinggi	Rendah.
Jarak tempat tinggal ke kantor	Dekat	Jauh
Umur/usia	Tua	Muda
Ketidakdisiplinan	Tinggi	Rendah
Pendidikan	Rendah	Tinggi

Gambar 9 dan Tabel 1 menunjukkan bahwa karyawan di cluster 0 merupakan sekumpulan pejabat/karyawan senior (level 4) dilihat dari biaya transportasi yang diberikan perusahaan, usia yang matang, dan pendidikannya yang rendah berbeda dengan yang ada di cluster 1 adalah sekumpulan karyawan milenial (level 1-3) yang usianya masih muda, pendidikannya tinggi, biaya transportasinya rendah tetapi memiliki kedisiplinan yang tinggi (ketidakdisiplinan rendah) sehingga dengan hasil analisa ini perusahaan wajib memberikan teguran/sanksi pemotongan biaya transportasi.

Setelah membagi menjadi 2 cluster dengan skenario kinerja yang “baik” dan “buruk” selanjutnya skenario ke-2 yaitu membagi 4 cluster dengan skenario kinerja “sangat baik”, “Baik”, “Cukup”, “Kurang” sebagai bagian dari laporan kinerja unit ke HRD sebagai parameter pemberian bonus dengan persentasi : Sangat Baik (A =10%), Baik (B = 8%), Cukup (C = 6%), Kurang (D = 4%), sebagaimana ditunjukkan pada Gambar 10 berikut.

**Gambar 10.** Grafik hasil kalsterisasi 4 cluster dan 5 atribut centroid

Berdasarkan Gambar 10 dengan mengubah nilai K pada parameter menjadi 4 cluster maka data akan diklasterisasikan menjadi 4 kategori, sehingga bisa di Analisa sebagaimana Tabel 2.

**Tabel 2.** Analisa kinerja berdasarkan atribut dan 4 cluster.

Atribut	Cluster			
	0 (40 karyawan 28-50 <sup>th</sup> )	1 (98 karyawan 28-41 <sup>th</sup> )	2 (382 karyawan 27-49 <sup>th</sup> )	3 (220 karyawan 34-58 <sup>th</sup> )
Biaya Transportasi	Tinggi	Rendah	Sedang	Sangat rendah
Jarak tempat tinggal ke kantor	Sedang	Dekat	Jauh	Sangat Dekat
Umur/usia	Sedang	Sangat Muda	Muda	Tua
Ketidakdisiplinan	Tinggi	Rendah	Rendah	Rendah
Pendidikan	Sedang	Tinggi	Sangat rendah	Rendah

Berdasarkan Gambar 10 dan dari data tabel 2 di atas cluster 1,2 dan 3 telah membuktikan secara teoritis proses kerja algoritma *K-Means* clustering yang merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama (pada atribut ketidaksiplinan) dikelompokkan dalam satu cluster yang sama sebagaimana dilihat pada Tabel 2 pada saat cluster dibagi menjadi 2 dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster yang lain. Metode *K-Means* Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

## 5. KESIMPULAN

Rekognisi karyawan menggunakan metode *K-Means* clustering dapat digunakan secara akurat dalam pengambilan keputusan menggunakan 4 klaster dan 5 atribut. Dua skenario yaitu skenario pertama dimana 0 dengan kinerja baik dan 1 dengan kinerja buruk dan skenario kedua yaitu cluster yang dibagi menjadi 4 dimana 0 dengan kinerja sangat baik, 1 kinerja baik, 2 kinerja cukup, 3 kinerja kurang dapat disimpulkan bahwa *K-Means* clustering mengelompokkan data yang ada ke dalam beberapa kelompok dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain sehingga sangat akurat dalam penarikan kesimpulan.

## DAFTAR PUSTAKA

- [1] R. J. S. Andrea Martiniano, Ricardo Pinto Ferreira, "Absenteeism at work Data Set," *National science foundation, Rexa.info*, 2018. .
- [2] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017.
- [3] A. Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016.
- [4] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," pp. 394–403, 2019.
- [5] W. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.
- [6] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020.
- [7] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," no. 1, pp. 26–32.
- [8] C. Wu *et al.*, "K-Means Clustering Algorithm and Its Simulation Based on Distributed Computing Platform," *Complexity*, vol. 2021, 2021.
- [9] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kaje Pekalongan," *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.
- [10] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019.