

***K-Means Clustering* Dalam Penerimaan Karyawan Baru**

Anggun Setiadi¹, Erma Delima Sikumbang^{2*}

¹ Sistem Informasi; STMIK Nusa Mandiri; Jl. Damai No.8 Warung Jati Barat (Margasatwa) Jakarta Selatan, 021-78839513; e-mail: anggunsetiadi99@gmail.com

² Sistem Informasi Akuntansi; Universitas Bina Sarana Informatika; Jl. Kramat Raya No. 98, Jakarta Pusat, 021- 21231170; e-mail: erma@bsi.ac.id

* Korespondensi: e-mail: erma@bsi.ac.id

Diterima: 23 April 2020 ; Review: 27 April 2020; Disetujui: 08 Mei 2020

Cara sitasi: Setiadi A, Sikumbang ED. 2020. *K-Means Clustering* Dalam Penerimaan Karyawan Baru. *Informatics for Educators and Professionals*. Vol.4 (2): 103-112.

Abstrak: Dalam penerimaan karyawan baru sulitnya bagian SDM PT. Erdikha Elit Sekuritas dalam mengelompokkan data-data karyawan baru dan tidak adanya sistem tes dalam pemilihan karyawan baru. Metode K-Means Clustering adalah salah satu metode cluster analysis non hirarki yang berusaha untuk mengelompokkan data-data yang ada satu atau lebih cluster atau kelompok, oleh karena itu metode ini sangat cocok digunakan untuk mengatasi permasalahan dalam mengelompokkan data-data calon karyawan baru dan mengimplementasikan menggunakan software RapidMiner dengan hasil penelitian 0,125% untuk cluster 1 yang berjumlah 2 data karyawan baru, 0,125% untuk cluster 2 yang berjumlah 2 data karyawan baru, dan 0,750% untuk cluster 3 yang berjumlah 12 data karyawan baru. Strategi pemilihan karyawan baru nantinya akan mengikuti cluster yang terbentuk berdasarkan data yang paling banyak diantara 3 cluster yang ada, yaitu di cluster ke- 3, karena dengan data cluster yang paling banyaklah yang lebih banyak memenuhi kriteria.

Kata kunci: *K-Means Clustering*, Penerimaan Karyawan Baru

Abstract: *In the case of hiring new employees, the difficulty of the HR department of PT. Erdikha Elit Sekuritas in classifying new employee data and the absence of a test system in the selection of new employees. K-Means Clustering method is a non-hierarchical cluster analysis method that seeks to group existing data into one or more clusters or groups, therefore this method is very suitable to be used to overcome problems in grouping data on prospective new employees and implements using RapidMiner software with research results of 0.125% for cluster 1 which amounts to 2 new employee data, 0.125% for cluster 2 which amounts to 2 new employee data, and 0.750% for cluster 3 which amounts to 12 new employee data. The new employee selection strategy will follow the clusters formed based on the most data among the 3 existing clusters, namely in the 3rd cluster, because with the most data clusters that meet more the required criteria.*

Keywords: *Acceptance of new employees, K-Means Clustering.*

1. Pendahuluan

Karyawan merupakan salah satu produksi yang terpenting dalam suatu perusahaan, tanpa karyawan perusahaan sulit mencapai tujuan, mereka dapat menentukan maju mudurnya sebuah perusahaan, dengan memiliki tenaga-tenaga kerja yang terampil dan motivasi yang tinggi sebuah perusahaan sudah mempunyai aset yang sangat mahal. Setiap perusahaan tentu memiliki tujuan yang ingin di capai, tujuan perusahaan pastilah mencapai laba yang optimal dalam jangka panjang sehingga kelangsungan hidup dari perusahaan tersebut dapat terjamin.

Seorang HRD (*Human Resource Departement*) Perusahaan untuk menentukan karyawan terbaik yang akan di pilih menjadi bagian dari perusahaan. Mulai dari verifikasi surat lamaran , sesi wawancara, pemeriksaan latar belakang serta referensi calon karyawan dari tempat kerja calon pelamar sebelumnya. Sulitnya PT. Erdikha Elit Sekuritas dalam mengelompokkan data calon karyawan baru dan tidak adanya sistem tes dalam perusahaan ini maka memanfaatkan teknologi informasi untuk mempermudah kinerja Staff HRD pada PT. Erdikha Elit Sekuritas untuk menentukan pilihan karyawan yang terbaik yang akan terpilih dan di tetapkan sebagai karyawan baru di PT. Erdikha Elit Sekuritas.

K-Means Clustering adalah, K dimaksudkan sebagai konstanta jumlah cluster yang diinginkan, Means dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai cluster, sehingga 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 [1].

Analisis Pengelompokan atau *Clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain. Potensi clustering adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola [2].

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases* (KDD) [3].

Data mining berisi pencarian trend atau pola yang diinginkan dalam *database* besar untuk membantu pengambilan keputusan diwaktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisis data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya [4].

Data mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang berbentuk tabulasi, seperti yang banyak diimplementasikan dalam teknologi manajemen basis data relasional [5].

Untuk menciptakan sumber daya manusia yang handal membutuhkan pengelolaan yang baik agar Kinerja Karyawan lebih optimal. Pencapaian tujuan perusahaan dipengaruhi oleh Kinerja Karyawan perusahaan itu sendiri. Maka dari itu perusahaan membutuhkan sumber daya manusia yang berpotensi dan berkualitas, baik dari segi pemimpin maupun karyawan [6]. Beberapa penulis terdahulu telah menerapkan K-Means Clustering sebagai penelitian dalam pengelompokan data antara lain Sibarani and Chafi (2018) K-Means merupakan salah satu metode pengelompokan data non-hirarki yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data yang ada ke dalam kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Penelitian ini bertujuan untuk menghasilkan suatu analisa yang dapat mengcluster atau mengelompokan data mahasiswa wisudawan tahun 2016 dan 2017 untuk membuat promosi menjadi lebih tertarget dan tepat sasaran. Penelitian ini mengklaster data mahasiswa lulusan dengan atribut alamat, jurusan dan IPK kedalam tiga Cluster berdasarkan jarak (Euclidean). Data yang diolah adalah data wisudawan tahun 2016-2017 [7]. Chasanah and Widiyono (2017) Penerimaan mahasiswa baru di STMIK Widya Pratama Pekalongan merupakan kegiatan setiap tahun. Setiap pelaksanaan terdapat beberapa kendala dan permasalahan. Masalah sejak tiga tahun yang lalu adalah penurunan jumlah pendaftar calon mahasiswa baru. Beberapa evaluasi dilakukan untuk meningkatkan jumlah pendaftar mahasiswa baru. Permasalahan yang lain antara lain kurangnya peminat program studi tertentu, sehingga pendaftar di program studi Akuntansi Komputer sedikit. Berdasarkan data pendaftaran siswa baru dalam penelitian ini akan dianalisa dengan metode pengelompokan / klustering menggunakan algoritma K-Means untuk mendapatkan informasi-informasi sesuai kelompok-kelompok tertentu dari data pendaftaran

tahun sebelumnya. Hasil klustering akan digunakan untuk menentukan strategi promosi penerimaan mahasiswa baru STMIK Widya Pratama Pekalongan [8].

Anggara et al (2016) Clustering merupakan salah satu metode Data Mining yang bersifat tanpa arahan (unsupervised). Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu hierarchical (hirarki) data clustering dan non-hierarchical (non hirarki) data clustering. K-Means Clustering merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Salah satu ciri clustering yang baik atau optimal adalah jika menghasilkan cluster yang berisi data dengan tingkat kemiripan (similarity) yang tinggi pada cluster yang sama dan tingkat kemiripan rendah pada cluster yang berbeda. Distance measure digunakan untuk mengukur kemiripan data dalam suatu cluster. Hasil dari proses clustering akan menghasilkan hasil yang berbeda apabila distance measure yang digunakan berbeda [9].

Tujuan penelitian ini adalah membantu perusahaan dalam menentukan pemilihan karyawan dengan menggunakan K-Means Clustering.

2. Metode Penelitian

Adapun Pengumpulan data yang dilakukan sebagai berikut: a. Observasi yaitu melakukan observasi di PT. Erdikha Elit Sekuritas yaitu melihat dan mengamati data-data calon karyawan baru yang ada di perusahaan. b. Wawancara yaitu mengadakan tanya jawab secara langsung kepada direktur PT. Erdikha Elit Sekuritas. c. Dokumentasi data calon karyawan baru di PT. Erdikha Elit Sekuritas selama bulan januari-desember 2018 yang nantinya akan diolah. Berikut kerangka penelitian yang dilakukan dijelaskan pada gambar 1.



Sumber: Hasil Penelitian (2019)

Gambar 1. Kerangka Pemikiran

Berdasarkan Gambar 1 kerangka penelitian yang digunakan mulai dari identifikasi masalah, pengumpulan data karyawan baru, Pengolahan data dengan metode Clustering, Evaluasi hasil penelitian dan kesimpulan hasil penelitian. Dalam pemilihan karyawan baru pada PT. Erdikha Elit Sekuritas ini menggunakan 3 variabel yaitu Pendidikan, Bahasa Asing dan IPK.

Untuk menghitung jarak terdekat, pada penelitian ini digunakan L_2 (Euclidian) distance space, jarak antara dua titik hitung dengan rumus sebagai berikut [9]:

$$DL_2(x_1, x_2) = \|x_1 - x_2\|_2 = \sqrt{\sum_{k=1}^p (x_{2j} - x_{1j})^2} \dots \dots \dots (1)$$

Keterangan:

P = Dimensi Data

X1 = Posisi titik 1

X2 = Posisi titik 2

Jarak yang terpendek antara centroid dengan dokumen menentukan posisi cluster suatu dokumen. Adapun rumus iterasi lainnya didefinisikan sebagai berikut [10]:

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \dots \dots \dots (2)$$

Keterangan:

X1 = Nilai data record ke-1

X2 = Nilai data record ke-2

3. Hasil dan Pembahasan

Data karyawan baru pada tahun 2018 yang dikelompokkan berdasarkan variabel-variabel yang ada pada data calon karyawan baru.

Tabel 1. Data Karyawan Baru

No.	NAMA KARYAWAN	PENDIDIKAN	BAHASA ASING	IPK
1	IVAN KASULTHAN	3	2	3,12
2	HAFIDH QARAZIA BARLY	3	2	3,14
3	ROWLINARI NATALIA NAPITUPULU	3	2	3,52
4	FITHRIA AL-AGHNI	3	2	3,44
5	SATRIA DUTA JULIANTO	2	1	2,76
6	RISKA HERNAWATI	3	2	3,17
7	MUHAMMAD DEWA BAGASKORO	3	2	3,09
8	DANIEL ANDREAS SITOANG	3	2	3,21
9	NURUL WAHYUNINGSIH	3	2	3,21
10	EDO ARDIANSYAH	3	2	3,43
11	JUNAIDY MUZHARY	3	2	3,32
12	M. ALIF AKBAR ADITAMA	1	1	3,18
13	THERESIA FRANSISCA	2	2	3,55
14	DEDE SUMARJONO	3	1	3,47
15	RUDI SAEFUDIN	3	2	3,27
16	MUHAMMAD SYAMSUL ARID	1	2	3,39

Sumber: Hasil Penelitian (2019)

Berdasarkan tabel 1 diatas data karyawan sebanyak 16 karyawan yang dilengkapi dengan variable pendidikan, Bahasa asing dan ipk.

Tabel 2. Data Pendidikan

PILIHAN	KETERANGAN
1	D4
2	D3
3	S1

Sumber: Hasil Penelitian (2019)

Berdasarkan tabel 2 diatas untuk pilihan 1 yaitu pendidikan D4, pilihan 2 pendidikan Diploma dan pilihan 3 pendidikan Sarjana.

Tabel 3. Nilai Bahasa Asing

PILIHAN	KETERANGAN
1	Tidak
2	Ya

Sumber: Hasil Penelitian (2019)

Berdasarkan table 3 diatas untuk pilihan 1 yaitu tidak menguasai Bahasa asing dan pilihan 2 menguasai bahasa asing. Tahapan analisis data dengan algoritma *k-means* memiliki langkah-langkah yang telah ditetapkan yaitu:

3.1. Menentukan Jumlah Cluster

Penelitian ini menetapkan 3 *Cluster* yang dipilih secara random pada variabel Pendidikan, Bahasa asing, dan IPK. Data yang dihasilkan dapat dilihat sebagai berikut :

Tabel 4. *Cluster* Awal

Cluster	Pendidikan	Bahasa Asing	IPK
C1	3	2	3,09
C2	3	2	3,21
C3	3	2	3,14

Sumber: Hasil Penelitian (2019)

Berdasarkan table 4 diatas dihasilkan data yaitu cluster 1 IPK 3,09, cluster 2 IPK 3,21 dan cluster 3 IPK 3,14.

3.2. Menghitung jarak tiap data ke pusat cluster

Penelitian ini menggunakan pengukuran jarak pada ruang jarak (*distance space*) Euclidean dengan menggunakan rumus[9] :

$$D_{i_2}(x_2, x_1) = \|X_2 - X_1\|^2$$

$$= \sqrt{(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 + (x_{23} - x_{13})^2} \dots\dots\dots(3)$$

Karena variabelnya ada 3 maka perhitungan x dilakukan sebanyak 3 kali sebagai perhitungan masing-masing variabel. Berikut ini merupakan contoh perhitungan jarak pada data ke 1 sampai data ke 16 sebagai berikut :

Contoh perhitungan dari data ke-1 ke pusat *cluster* 1

$$d(1,1) = \sqrt{(3 - 3)^2 + (2 - 2)^2 + (3,12 - 3,09)^2}$$

$$d(1,1) = \sqrt{(0)^2 + (0)^2 + (0,03)^2}$$

$$d(1,1) = \sqrt{0,0009}$$

$$d(1,1) = 0,03$$

Lanjutkan perhitungan data ke 2 sampai data ke 16

Dari data ke-1 ke pusat *cluster* 2

$$d(1,2) = \sqrt{(3 - 3)^2 + (2 - 2)^2 + (3,12 - 3,21)^2}$$

$$d(1,2) = \sqrt{(0)^2 + (0)^2 + (-0,09)^2}$$

$$d(1,2) = \sqrt{0,0081}$$

$$d(1,2) = 0,09$$

Lanjutkan perhitungan data ke 2 sampai data ke 16

Dari data ke-1 ke pusat *cluster* 3

$$d(1,3) = \sqrt{(3 - 3)^2 + (2 - 2)^2 + (3,12 - 3,14)^2}$$

$$d(1,3) = \sqrt{(0)^2 + (0)^2 + (-0,02)^2}$$

$$d(1,3) = \sqrt{0,0004}$$

$$d(1,3) = 0,02$$

Lanjutkan perhitungan data ke 2 sampai data ke 16

Setelah dilakukan perhitungan, maka hasil perhitungan selengkapnya dapat dilihat pada tabel berikut

Tabel 5. Hasil Perhitungan Jarak Data Pada Iterasi 1

Data Ke-i	Nilai Cluster			Cluster		
	1	2	3	1	2	3
1	0,03	0,09	0,02			*
2	0,05	0,07	0			*
3	0,43	0,31	0,38		*	
4	0,35	0,23	0,3		*	
5	1,4522	1,4840	1,4643	*		
6	0,08	0,04	0,03			*
7	0	0,12	0,05	*		
8	0,12	0	0,07		*	
9	0,12	0	0,07		*	
10	0,34	0,22	0,29		*	
11	0,23	0,11	0,18		*	
12	2,2378	2,2362	2,2364		*	
13	1,1007	1,0562	1,0807		*	
14	1,0697	1,0332	1,0530		*	
15	0,18	0,06	0,13		*	
16	2,0223	2,0080	2,0155		*	

Sumber: Hasil Penelitian (2019)

Berdasarkan hasil perhitungan pada table 5 diatas, setiap data akan menjadi anggota suatu *cluster* yang memiliki jarak terdekat (nilai yang terkecil) dari pusat *clusternya*. Misal di data pertama nilai yang terkecil ada pada *cluster* ke-3 dengan memperoleh hasil 0,02.

3.3. Mendapatkan Pusat *Cluster* Baru

Setelah data dihitung dan ditempatkan pada masing-masing *cluster* terdekat, maka selanjutnya menghitung kembali data untuk mendapatkan pusat *cluster* yang baru yakni berdasarkan nilai rata-rata dari masing-masing variabel, pada masing-masing *cluster*. Perhitungannya adalah sebagai berikut :

Rata-rata pendidikan pada cluster 1 :

Data ke 5 dan 7

$$\begin{aligned}
 &= \frac{2 + 3}{2} \\
 &= \frac{5}{2} \\
 &= 2,5
 \end{aligned}$$

Rata-rata bahasa asing pada cluster 1 :

Data ke 5 dan 7

$$\begin{aligned}
 &= \frac{1 + 2}{2} \\
 &= \frac{3}{2} \\
 &= 1,5
 \end{aligned}$$

Rata-rata IPK pada cluster 1 :

Data ke 5 dan 7

$$\begin{aligned}
 &= \frac{2,76 + 3,09}{2} \\
 &= \frac{5,85}{2}
 \end{aligned}$$

$$= 2,925$$

Rata-rata pendidikan pada cluster 2 :

Data ke 3, 4, 8, 9, 10, 11, 12, 13, 14, 15 dan 16

$$\begin{aligned} &= \frac{3 + 3 + 3 + 3 + 3 + 3 + 1 + 2 + 3 + 3 + 1}{11} \\ &= \frac{28}{11} \\ &= 2,545 \end{aligned}$$

Rata-rata bahasa asing pada cluster 2 :

Data ke 3, 4, 8, 9, 10, 11, 12, 13, 14, 15 dan 16

$$\begin{aligned} &= \frac{2 + 2 + 2 + 2 + 2 + 2 + 1 + 2 + 1 + 2 + 2}{11} \\ &= \frac{20}{11} \\ &= 1,818 \end{aligned}$$

Rata-rata IPK pada cluster 2 :

Data ke 3, 4, 8, 9, 10, 11, 12, 13, 14, 15 dan 16

$$\begin{aligned} &= \frac{3,52 + 3,44 + 3,21 + 3,21 + 3,43 + 3,22 + 3,18 + 3,55 + 3,47 + 3,27 + 3,39}{11} \\ &= \frac{36,99}{11} \\ &= 3,362 \end{aligned}$$

Rata-rata pendidikan pada cluster 3 :

Data ke 1, 2 dan 6

$$\begin{aligned} &= \frac{3 + 3 + 3}{3} \\ &= \frac{9}{3} \\ &= 3 \end{aligned}$$

Rata-rata bahasa asing pada cluster 3 :

Data ke 1, 2 dan 6

$$\begin{aligned} &= \frac{2 + 2 + 2}{3} \\ &= \frac{6}{3} \\ &= 2 \end{aligned}$$

Rata-rata IPK pada cluster 3 :

Data ke 1, 2 dan 6

$$\begin{aligned} &= \frac{3,12 + 3,14 + 3,17}{3} \\ &= \frac{9,43}{3} \\ &= 3,143 \end{aligned}$$

Berdasarkan hasil perhitungan dapat disimpulkan bahwa cluster 1 memiliki rata-rata ipk 2,925, cluster 2 memiliki rata-rata 3,362 dan cluster 3 memiliki rata-rata ipk 3,143.

Data Ke-i	Nilai Cluster			Cluster		
	1	2	3	1	2	3
1	0,733502	0,54643	0,023333			*
2	0,73907	0,537844	0,003333			*
3	0,924135	0,514202	0,376667			*
4	0,874771	0,495621	0,296667			*
5	0,726103	1,153353	1,465245	*		
6	0,748348	0,52613	0,026667			*
7	0,726103	0,560401	0,053333			*
8	0,762381	0,51283	0,066667			*
9	0,762381	0,51283	0,066667			*
10	0,868922	0,49416	0,286667			*
11	0,809954	0,491421	0,176667			*
12	1,60157	1,758193	2,236369	*		
13	0,943729	0,60469	1,079527		*	
14	0,892763	0,942094	1,052003	*		
15	0,786781	0,498264	0,126667			*
16	1,648097	1,556352	2,015154		*	

Sumber: Hasil Penelitian (2019)

Gambar 2. Hasil Perhitungan Jarak Data Pada Iterasi 2

Berdasarkan hasil perhitungan pada gambar 2 diatas, setiap data akan menjadi anggota suatu cluster yang memiliki jarak terdekat (nilai yang terkecil) dari pusat clusternya. Misal di data pertama nilai yang terkecil ada pada cluster ke-3 dengan memperoleh hasil 0,00

Karena centroid baru yang digunakan belum konvergen maka iterasi harus dilanjutkan. Pada penelitian ini iterasi terjadi sebanyak 4 kali karena kondisi cluster sudah mencapai konvergen sehingga proses iterasi berhenti. Hasil akhir dari perhitungan dapat dilihat pada gambar dibawah ini.

Data Ke-i	Nilai Cluster			Cluster		
	1	2	3	1	2	3
1	1,809005	1,540292	0,182622			*
2	1,810773	1,535871	0,165078			*
3	1,884808	1,500833	0,251696			*
4	1,863035	1,5003	0,178187			*
5	0,54231	1,324424	1,453714	*		
6	1,813836	1,529706	0,140002			*
7	1,806765	1,547385	0,209763			*
8	1,818681	1,522367	0,110457			*
9	1,818681	1,522367	0,110457			*
10	1,860538	1,500533	0,169413			*
11	1,836437	1,507481	0,091382			*
12	0,54231	1,155032	2,20245	*		
13	1,259524	0,50636	1,038509		*	
14	1,581139	1,802776	0,935647			*
15	1,827567	1,513275	0,084265			*
16	1,19432	0,50636	2,00462		*	

Sumber: Hasil Penelitian (2019)

Gambar 3. Hasil Perhitungan Jarak Data Pada Iterasi 4

Berdasarkan hasil perhitungan pada gambar 3 diatas, pada iterasi ke 3 dan iterasi ke 4 tidak ada perubahan maka perhiungan berhenti pada iterasi ke 4. Pada Gambar 4 adalah Hasil Perhitungan Manual dengan *RapidMiner*.

Nama Karyawan	Manual	RapidMiner
Satria Duta Julianto	0,125%	0,125%
M. Alif Akbar A ditama		
Theresia Fransisca Muhammad Syamsul Arid Ivan Kasulthan	0,125%	0,125%
Hafidh Qarazia Barly	0,750%	0.750%
Rowlinari Natalia Napitupulu		
Fithria Al-Aghni		
Riska Hernawati		
Muhammad Dewa Bagaskoro		
Daniel Andreas Sitohang		
Nurul Wahyuningsih		
Edo Ardiansyah		
Junaidy Muzhary		
Dede Sumarjono		
Rudi Saefudin		

Sumber: Hasil Penelitian (2019)

Gambar 4. Perbandingan Hasil Perhitungan Manual Dengan *RapidMiner*

Berdasarkan perhitungan manual dan rapidminer dapat disimpulkan bahwa hasil dari perhitungan diatas yaitu mendapatkan data untuk *cluster* 1 berjumlah 2 data karyawan, *cluster* 2 berjumlah 2 data karyawan dan *cluster* 3 mendapatkan 12 data karyawan. Jadi untuk perhitungan manual dan *RapidMiner* hasil datanya sama dan tidak ada yang berubah.

Row No.	NAMA KARY...	cluster	PENDIDIKAN	BAHASA ASI...	IPK
1	IVAN KASULT...	cluster_2	3	2	3.120
2	HAFIDH QAR...	cluster_2	3	2	3.140
3	ROWLINARI ...	cluster_2	3	2	3.520
4	FITHRIA AL-A...	cluster_2	3	2	3.440
5	SATRIA DUT...	cluster_0	2	1	2.760
6	RISKA HERN...	cluster_2	3	2	3.170
7	MUHAMMAD ...	cluster_2	3	2	3.090
8	DANIEL AND...	cluster_2	3	2	3.210
9	NURUL WAH...	cluster_2	3	2	3.210
10	EDO ARDIAN...	cluster_2	3	2	3.430
11	JUNAIDY MU...	cluster_2	3	2	3.320
12	M. ALIF AKBA...	cluster_0	1	1	3.180
13	THERESIA F...	cluster_1	2	2	3.550
14	DEDE SUMA...	cluster_2	3	1	3.470
15	RUDI SAEFU...	cluster_2	3	2	3.270
16	MUHAMMAD ...	cluster_1	1	2	3.390

Sumber: Hasil Penelitian (2019)

Gambar 5. Hasil Data Cluster dengan *RapidMiner*

4. Kesimpulan

Dengan menggunakan algoritma k-means perusahaan dapat mengelompokkan data karyawan baru berdasarkan variabel seperti pendidikan, bahasa asing dan IPK agar dalam perhitungan menjadi lebih akurat. Setelah dilakukan pengelompokkan data hasil akurasi yang didapatkan setelah dihitung adalah: 0,125% untuk cluster 1 yang berjumlah 2 data karyawan baru, 0,125% untuk cluster 2 yang berjumlah 2 data karyawan baru, 0,750% untuk cluster 3 yang berjumlah 12 data karyawan baru. Strategi pemilihan karyawan baru nantinya akan mengikuti cluster yang terbentuk berdasarkan data yang paling banyak diantara 3 cluster yang ada, yaitu di cluster ke-3, karena dengan data cluster ke-3 yang paling banyak memenuhi kriteria yang dibutuhkan, yang nantinya akan diproses lebih lanjut oleh HRD yaitu Ivan Kasulthan, Hafidh Qarazia Barly, Rowlinari Natalia Napitupulu, Fithria Al-Aghni, Riska Hernawati, Muhammad Dewa Bagaskoro, Daniel Andreas Sitohang, Nurul Wahyuningsih, Edo Ardiansyah, Junaidy Muzhary, Dede Sumarjono, Rudi Saefudin.

Referensi

- [1] N. Fina and D. Surya, "Seminar Nasional Teknologi Informasi dan Multimedia 2015 Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama)," pp. 6–8, 2015.
- [2] S. W. Riyani, W. Anjar, and W. P. Agus, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018.
- [3] V. T. Retno, *Data Mining Teori dan Aplikasi Rapidminer*. Surakarta: Media, Gava, 2017.
- [4] H. A. Fajar, *Data Mining*. Surabaya: ANDI OFFSET, 2013.
- [5] S. M. Amril and P. Adam, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV. Kekata Group, 2018.
- [6] W. Rita, "Pengaruh Pengawasan Internal Dan Kepemimpinan Terhadap Kinerja Karyawan (Studi Kasus Di Politeknik Lp3I Jakarta)," *Maj. Ilm. Bijak*, vol. 15, no. 1, pp. 77–90, 2018.
- [7] S. Riama and Chafid, "Algorithma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indonesia [Algorithma K-Means Clustering Strategy Marketing Admission Universitas Satya Negara Indonesia]," *Semin. Nas. Cendekiawan ke 4*, no. 2, pp. 685–690, 2018.
- [8] T. Tria and W. Chasanah, "Penentuan Strategi Promosi Penerimaan Mahasiswa Baru Dengan Algoritma Clustering K-Means," vol. XII, no. 25, pp. 39–44, 2017.
- [9] A. Mario, S. Herry, and H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness," *J. Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 1–6, 2016.
- [10] Y. Fitri, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018.