

Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan *Algoritma K-Means Clustering* (Studi Kasus Di Pt Indomex Dwijaya Lestari)

Novi Yona Sidratul Munti¹, Gunadi Widi Nurcahyo², Julius Santony³

¹ Jurusan Informatika FST UPTT

Jln. Tuanku Tambusai No.23 Bangkinang 28412 INDONESIA

¹noviyona@universitaspahlawan.ac.id

^{2,3} Universitas Putra Indonesia YPTK Padang

Jln Lubuk Begalung Padang INDONESIA

² gunadiwidinurcahyo@gmail.com

³ Julius_santony@yahoo.com

Intisari— Perusahaan adalah sebuah unit kegiatan produksi yang mengolah sumber daya ekonomi untuk menyediakan barang dan jasa bagi masyarakat dengan tujuan memperoleh keuntungan dan memuaskan kebutuhan masyarakat. Pencarian gaji karyawan saat ini masih dilakukan secara manual, sehingga masih banyak ditemukan kesimpangan dalam penerimaan gaji, sehingga menyebabkan beberapa karyawan mengeluh saat penerimaan gaji, untuk itu diperlukan teknik dalam menentukan gaji karyawan Tetap dan karyawan Kontrak. Salah satu teknik yang dapat digunakan adalah Menggunakan *Algoritma K-Means Clustering*.

Kata kunci— *Clustering Algoritma K-Means*, *Rapid Miner*.

Abstract— The company is a unit of production activities that cultivate economic resources to provide goods and services for society with the aim of gaining profit and satisfy the needs of the community. Within the company we need to determine the classification of a worker for the purpose of assessing profesionalisme employess, in order to improve the quality of workers in a company. Search employee's salary is still done manually, so there are still many in receipt of a salary, so that led to some employees complained upon receipt of salary, it is necessary for the technique in determining the employee's salary permanent and Contract employees. One technique that can be used are Using *K-Means Clustering Algorithm*.

Keywords— *Salary, Data Mining, Clustering, Rapid Miner*.

I. PENDAHULUAN

Pekerjaan dalam arti luas adalah aktivitas utama yang dilakukan oleh manusia. Dalam arti sempit, istilah pekerjaan digunakan untuk suatu tugas atau kerja yang menghasilkan uang bagi seseorang. Dalam pembicaraan sehari-hari istilah ini sering dianggap sinonim dengan profesi.

Di dalam Perusahaan kita perlu untuk menentukan klasifikasi seorang pekerja dengan tujuan untuk menilai Profesionalisme karyawan, guna untuk meningkatkan mutu para pekerja dalam sebuah perusahaan.

Berdasarkan penjelasan di atas, maka penulis menyimpulkan bahwa penentuan gaji karyawan merupakan proses pengalokasian sumber daya yang ada sesuai dengan fungsinya untuk mencapai suatu tujuan. Dengan demikian kita dapat menentukan perbandingan gaji karyawan Tetap dan karyawan Kontrak.

Pencarian gaji karyawan saat ini masih dilakukan secara manual, sehingga masih banyak ditemukan kesimpangan dalam penerimaan gaji, sehingga menyebabkan beberapa karyawan mengeluh saat penerimaan gaji, untuk itu diperlukan teknik dalam menentukan gaji karyawan Tetap dan



karyawan Kontrak. Salah satu teknik yang dapat digunakan adalah Menggunakan Algoritma K-Means Clustering.

Metode clustering K-Means banyak digunakan untuk proses segmentasi, contohnya dalam proses segmentasi pelanggan sebuah supermarket yang dilakukan oleh Sangkar (2011). Teknik Data Mining juga diterapkan dalam penelitian sebelumnya untuk menentukan Survei yang berbeda dengan Algoritma Berdasarkan Dataset Tata Ruang (Maitry dan Vaghela, 2014). Data Mining dengan Sistem Terdistribusi dan hemat energi menggunakan Metode Clustering untuk hirarkis Wireless Sensor NetworksHuang , dkk (2013).

II. LANDASAN TEORI

A. Knowledge Discovery In Database (KDD).

Data Mining dapat dikatakan ekstraksi informasi atau pola yang penting atau sesuatu yang menarik dari data yang ada di Database berskala besar yang dikenal dengan Knowledge Discovery in Database (KDD). Proses ekstraksi dilakukan untuk menemukan pengetahuan baru yang menarik seperti pola, asosiasi, aturan (*rule*), perubahan, keganjilan dan struktur penting dari sejumlah data besar yang tersimpan di dalam database atau file informasi lainnya.

B. Definisi Data Mining

Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*mechine learning*) untuk menganalisis data dan mengekstraksi pengetahuan (*knowledge*) secara otomatis menurut Fajar Astuti Hermawati (Nurhayati, 2014).

C. Tahapan Data Mining.

Tahapan – tahapan yang dilakukan dalam proses Data Mining (Adi Sucipto, 2015) :

1. Data Cleaning, merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan.
2. Data Integration, merupakan proses penggabungan data ke dalam satu database di mana terkadang proses Data Mining yang akan dilakukan memerlukan data lebih dari satu database atau memerlukan pemrosesan menggunakan database lain.
3. Data Transpormation, merupakan pengubahan data atau penggabungan data yang dilakukan sebelum akan dilakukan proses Data Mining. Karena dalam proses Data Mining dalam beberapa metode membutuhkan format data yang khusus sebelum data dialokasikan.
4. Aplikasi Teknik Data Mining, merupakan salah satu bagian dari proses Data Mining.
5. Interpretasi / Evaluasi, merupakan pola yang menentukan hasil dari teknik Data Mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa memang benar-benar tercapai.
6. Knowledge / Pengetahuan, merupakan presentasi pola yang ditemukan untuk menghasilkan aksi tahap terakhir dari proses Data Mining dan bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat.

D. Defenisi Clustering.

Clustering adalah tanpa pengawasan mekanisme klasifikasi di mana satu set pola (data), biasanya multidimensi diklasifikasikan ke dalam kelompok seperti bahwa anggota satu kelompok yang sama sesuai dengan kriteria yang telah ditetapkan.

E. Algoritma K-Means

K-means merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi dataset kedalam beberapa kluster k . Algoritmanya cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan menurut Wu & Kumar (Widiarina,2015). Prinsip utama dari teknik ini adalah menyusun k buah partisi/pusat massa (centroid)/ratarata (*mean*) dari sekumpulandata. Algoritma K-means dimulai dengan pembentukan partisi kluster di awal kemudian secara iteratif partisi kluster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi kluster menurut Witten, Eibe, & Hall (Widiarina,2015).

III. METODOLOGI PENELITIAN

Kerangka kerja penelitian merupakan tahapan-tahapan yang dilakukan oleh penulis dalam menyelesaikan penelitian. Adapun kerangka kerja penelitian ini dengan menggunakan metode clustering Algoritma K-Means pada gambar



Gambar 1 Tahapan Penelitian

IV. HASIL DAN PEMBAHASAN

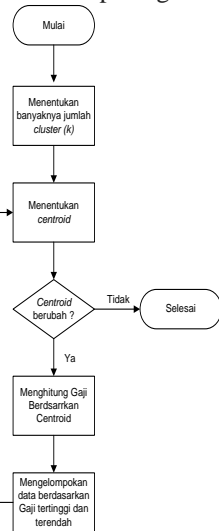
A. Analisa Data

Dalam mengumpulkan data dilakukan dengan cara melakukam wawancara serta mengambil data kepada administrasi kantor tersebut. Faktor-faktor yang mempengaruhi gaji karyawan banyak jenisnya, diantaranya faktor *Intern* dan faktor *Ekstern*. Namun pada penelitian ini faktor yang digunakan adalah faktor *Intern*. Faktor-faktor ini

nantinya akan dikelompokkan sehingga membantu dalam memahami data gaji karyawan. Penganalisaan dilakukan berdasarkan fakta dan data yang didapat di PT Indomex Dwijaya Lestari.

B. Analisa Clustering dengan Menggunakan Algoritma K-Means

Analisis cluster merupakan pengelompokan objek-objek data hanya berdasarkan pada informasi yang terdapat pada data, yang menjelaskan objek dan relasinya. Tahapan algoritma K-Means dapat dilihat pada gambar berikut:



Gambar 2 Diagram Proses Flowchart K-Means.

C. Penetapan Data.

Sebelum melakukan proses pengelompokan data Karyawan, terlebih dahulu tentukan data yang akan dilakukan pengolahan agar tujuan penelitian dapat tercapai..

D. Menentukan Atribut Yang Digunakan

Pada penelitian ini faktor yang mempengaruhi gaji karyawan tetap dan karyawan kontrak yaitu:

1. Bagian kerja karyawan
2. Status kerja karyawan
3. Gaji Pokok karyawan
4. Hari kerja karyawan
5. Total gaji karyawan

E. Pengkodean Data

Untuk membuktikan bahwa faktor rekap gaji karyawan dipengaruhi oleh faktor – faktor yang telah ditentukan seperti yang sudah dijelaskan sebelumnya makadibutuhkan suatu alat bantu untuk membuktikannya yaitu berupa rekap gaji karyawan dengan melihat total gaji karyawan. Adapun data yang diolah yaitu karyawan tetap sebanyak 32 orang, sedangkan karyawan kontrak 28 orang, pada tabel berikut adalah data yang mewakili saja, terlihat pada tabel berikut:

Tabel 4.1 :Sampel Data Awal Karyawan Tetap dan Karyawan Kontrak.

No	Nama	Bagian	Status	Gaji Pokok	Jumlah Hari Kerja	Yang Diterima
1	KUDI DAKMAWI	DIKERJE	TETAP	Rp. 3.000,000	25	Rp. 3.529,075
2	RIKO MAYEK	GENERAL MANAGER	TETAP	Rp. 2.500,000	25	Rp. 3.100,675
3	MAWARDI	MEKANIK	TETAP	Rp. 2.200,000	25	Rp. 2.715,075
4	MAHLAN	MEKANIK	TETAP	Rp. 2.200,000	25	Rp. 2.715,075
5	MULFARI	MANAGER CP/MD	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
6	MAWARDI K	MANAGER AMEK	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
7	ADHARI AGUSTI	MANAGER AKUNTANSI	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
8	JULIANO	MANAGER PEMASARAN	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
9	GESTIANI BULUOLO	PERSONALIA DAN UMUM	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
10	SARDUTI	PRODUKSI II PLANT	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
11	ANDHKA	PENGAWAS PRODUKSI	TETAP	Rp. 2.500,000	25	Rp. 3.015,075
12	YOLA	ADM PEMBELIAN	TETAP	Rp. 2.000,000	20	Rp. 2.560,075
13	MERI	ADM PENJUALAN	TETAP	Rp. 2.000,000	20	Rp. 2.560,075
14	KIMA	ADM PENJUALAN	TETAP	Rp. 2.000,000	20	Rp. 2.560,075
15	WULAN	ADM PENJUALAN	TETAP	Rp. 2.000,000	20	Rp. 2.560,075
16	FAUZI	SDM	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
17	RINA SALANGAN	KASIR	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
18	MERI KOWITA	ADM BANK	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
19	PERDIANTO	GUDANG	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
20	MAISARIL	KA GUDANG BARANG JADI	TETAP	Rp. 2.014,000	25	Rp. 2.523,355
21	VIARA Z	ADM HUTANG	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
22	KETHA	ADM GUDANG	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
23	OKY	ADM GUDANG	TETAP	Rp. 2.000,000	25	Rp. 2.560,075
24	WANU	SATPAM	TETAP	Rp. 2.000,000	25	Rp. 2.560,075

No	Nama	Bagian	Status	Gaji Pokok	Jumlah Hari Kerja	Yang Diterima
25	DODI	SATPAM	TETAP	Rp. 2.050,000	25	Rp. 2.560,075
26	OYONG	DRIVER	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
27	RICHI	DRIVER	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
28	RUDINI	DRIVER	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
29	RANISAT	DRIVER	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
30	DEFA	DRIVER	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
31	RIDI	DRIVER	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
32	ADRIION	SDM	TETAP	Rp. 2.014,000	20	Rp. 2.523,355
33	SLAMET	MEKANIK	KONTRAK	Rp. 1.800,725	25	Rp. 1.833,765
34	PUTRA	DRIVER	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
35	BADI	ANGGOTA GUDANG	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
36	VITRA	ANGGOTA GUDANG	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
37	HARIS	ANGGOTA GUDANG	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
38	ERI	SALES	KONTRAK	Rp. 1.800,725	18	Rp. 1.833,765
39	HANI	SALES	KONTRAK	Rp. 1.800,725	18	Rp. 1.833,765
40	HENDRA	SALES	KONTRAK	Rp. 1.800,725	18	Rp. 1.833,765
41	YOPI	SALES	KONTRAK	Rp. 1.800,725	18	Rp. 1.833,765
42	JAMRI	SATPAM	KONTRAK	Rp. 2.050,000	25	Rp. 2.088,025
43	CHAN	SATPAM	KONTRAK	Rp. 2.050,000	25	Rp. 2.088,025
44	ARIF	BOILER	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
45	DONI	BOILER	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
46	JHON	FORMULA	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
47	AGUS	FORMULA	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
48	SYAFRIAL	PACKING	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
49	ADHAR	PACKING	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
50	VINA	PACKING	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
51	DIKA	ANGGOTA PRODUKSI	KONTRAK	Rp. 2.050,000	28	Rp. 2.088,025
52	SISKA	ANGGOTA PRODUKSI	KONTRAK	Rp. 2.050,000	28	Rp. 2.088,025
53	MAYAR	SERVICE CLENING	KONTRAK	Rp. 1.800,725	25	Rp. 1.833,765
54	LASTRI	SERVICE CLENING	KONTRAK	Rp. 1.800,725	15	Rp. 1.833,765
55	MERI	SERVICE CLENING	KONTRAK	Rp. 1.800,725	25	Rp. 1.833,765
56	ADRIION	SERVICE CLENING	KONTRAK	Rp. 1.800,725	25	Rp. 1.833,765
57	YULENDRI ANI	SDM	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
58	ADRIION	SDM	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
59	MAISARIL	GUDANG BARANG JADI	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765
60	OKY	GUDANG BARANG JADI	KONTRAK	Rp. 1.800,725	20	Rp. 1.833,765

F. Transformasi Data

Transformasi data dilakukan karena jenis data tidak berupa numeric maka data harus ditransformasikan terlebih dahulu dengan cara melakukan frekuensi pada data yang terbanyak muncul dengan mengurutkan frekuensi tertinggi ke terendah dan lakukan inisial data

Tabel 4.1 Sampel data karyawan proses Clustering

No	Nama	Bagian	Status	Gaji Pokok	Jumlah Hari Kerja	Yang Diterima
1	RUDI DARMAWI	1	1	1	1	3
2	RIKO MAYER	2	1	2	1	3
3	MAWARDI	3	1	3	1	2
4	MASLAN	3	1	3	1	2
5	MULFAJRI	4	1	4	1	2
6	MAWARDIK	5	1	4	1	2
7	AZHARI AGUSTI	6	1	4	1	2
8	JULIANO	7	1	4	1	2
9	GESTIANI	8	1	4	1	2
10	SARDEWI	9	1	4	1	2
11	ANDHIKA	10	1	4	1	2
12	YOLA	11	1	5	2	2
13	MERI	12	1	5	2	2
14	RIMA	12	1	5	2	2
15	WULAN	12	1	5	2	2
16	FAUZ	13	1	5	1	2
17	RINA	14	1	5	1	2
18	MEISI KOSWITA	15	1	5	1	2
19	FERDIANTO	16	1	5	1	2
20	MAISARIL	17	1	6	1	2
21	VIARA Z	18	1	5	1	2
22	RETHA	19	1	5	1	2
23	OKY	19	1	5	1	2
24	WANDI	20	1	5	1	2
25	DODI	20	1	5	1	2
26	OYONG	21	1	6	2	2
27	RICHI	21	1	6	2	2

No	Nama	Bagian	Status	Gaji Pokok	Jumlah Hari Kerja	Yang Diterima
28	RUDSDI	21	1	6	2	2
29	RAHMAT	21	1	6	2	2
30	DEFIS	21	1	6	2	2
31	RUDI	21	1	6	2	2
32	ADRIZON	13	1	6	2	2
33	SLAMET	3	2	7	1	1
34	PUTRA	21	2	7	2	1
35	BADI	22	2	7	2	1
36	VITRA	22	2	7	2	1
37	HARIS	22	2	7	2	1
38	ERI	23	2	7	3	1
39	HADI	23	2	7	3	1
40	HENDRA	23	2	7	3	1
41	YOFI	23	2	7	3	1
42	JANGU	20	2	5	1	2
43	CHAN	20	2	5	1	2
44	ARIF	24	2	7	2	1
45	DONI	24	2	7	2	1
46	JHON	25	2	7	2	1
47	AGUS	26	2	7	2	1
48	SYAFRIZAL	26	2	7	2	1
49	ADHAR	26	2	7	2	1
50	VENA	26	2	7	2	1
51	DIKA	27	2	5	4	2
52	SISKA	27	2	5	4	2
53	MAYAR	28	2	7	1	1
54	LASTI	28	2	7	1	1
55	MERI	28	2	7	1	1
56	ADRIZON	28	2	7	1	1
57	YULENDRADI	13	2	7	2	1
58	ADRIZON	13	2	7	2	1
59	MAHARIL	29	2	7	2	1
60	OKY	29	2	7	2	1

G. Proses Clustering Menggunakan Algoritma K-Means

Data yang sudah ditetapkan akan dilakukan pengolahan data dengan proses *clustering* dengan menggunakan algoritma *K-Means* sehingga didapatkanlah hasil pengelompokan terbaik yang nantinya akan membantu dalam pemahaman data karyawan.

1. Penentuan jumlah cluster

Penentuan jumlah cluster dilakukan untuk mengetahui hasil dari total gaji karyawan. Maka dalam penelitian ini jumlah cluster yang digunakan adalah sebanyak 4 cluster ($k=4$),

sehingga nanti akan diketahui cluster terbaik dalam melakukan penelitian ini.

2. Menentukan centroid

Pusat awal cluster atau centroid ditentukan secara random atau acak, dimana nilai cluster 0 diambil dari baris ke-11, nilai cluster 1 pada baris ke-47. Berikut nilai titik pusat awal cluster (centroid) yang dapat dilihat pada tabel

4.3 Titik Pusat Awal Cluster:

NO	NAMA	BAGIAN	STATUS	GAJI POKOK	JUMLAH H.K	TOTAL
11	ANDHIKA	10	1	4	1	2
47	AGUS	25	2	7	2	1

Cluster 0 : (11 ; 10 ; 1 ; 4 ; 1 ; 2)

Cluster 1 : (47 ; 25 ; 2 ; 7 ; 2 ; 1)

3. Menentukan centroid

Pusat Menghitung jarak antara titik *centroid* dengan titik tiap objek dengan menggunakan *Euclidian Distance*. Ada pun penghitungan *centroid* awal secara manual. Perhitungannya adalah sebagai berikut:

Perhitungan dengan rumus :

$$D(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ik} - X_{jk})^2}$$

Perhitungan jarak dari data ke-1 terhadap pusat *cluster* :

$$\text{Centroid1} = \sqrt{(1-10)^2 + (1-1)^2 + (1-4)^2 + (1-1)^2 + (2-2)^2} = 90$$

$$\text{Centroid2} = \sqrt{(2-10)^2 + (1-1)^2 + (2-4)^2 + (1-1)^2 + (3-2)^2} = 69$$

$$\text{Centroid3} = \sqrt{(3-10)^2 + (1-1)^2 + (3-4)^2 + (1-1)^2 + (2-2)^2} = 50$$

$$\text{Centroid 4} = \sqrt{(3-10)^2 + (1-1)^2 + (3-4)^2 + (1-1)^2 + (2-2)^2} = 50$$

$$\text{Centroid 5} = \sqrt{(4-10)^2 + (1-1)^2 + (4-4)^2 + (1-1)^2 + (2-2)^2} = 36$$

Perhitungan jarak dari data ke-2 terhadap pusat *cluster* :

$$\text{Centroid1} = \sqrt{(1-25)^2 + (1-2)^2 + (1-7)^2 + (1-2)^2 + (2-1)^2} = 615$$

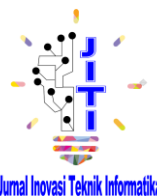
$$\text{Centroid2} = \sqrt{(2-25)^2 + (1-2)^2 + (2-7)^2 + (1-2)^2 + (3-1)^2} = 560$$

$$\text{Centroid3} = \sqrt{(3-25)^2 + (1-2)^2 + (3-7)^2 + (1-2)^2 + (2-1)^2} = 503$$

$$\text{Centroid4} = \sqrt{(3-25)^2 + (1-2)^2 + (3-7)^2 + (1-2)^2 + (2-1)^2} = 503$$

$$\text{Centroid5} = \sqrt{(4-25)^2 + (1-2)^2 + (4-7)^2 + (1-2)^2 + (2-1)^2} = 453$$

Alokasikan masing-masing objek ke *centroid* terdekat, untuk melakukan mengalokasikan objek ke dalam masing-masing *cluster* dengan cara mengelompokkan berdasarkan jarak minimum objek ke pusat *cluster* dengan memberikan kode "1" jika hasil *cluster* mendekati nol. Perhitungan dilakukan terus sampai data ke-60 terhadap pusat *cluster*. Setelah dilakukan proses perhitungan maka akan didapatkan data seperti Tabel berikut:



Tabel 2.4 Hasil Perhitungan Gaji dan Pengelompokan Data Iterasi ke-1

No	NAMA	Bagian	STATUS	Gaji Pokok	JUMLAH (Yang Diterima)	centroid 1				
						10	1	4	1	1
1	RUDI DARMAWI	1	1	1	1	2				
2	RINDI MAHER	2	1	2	1	3				
3	MAMARADI	3	1	3	1	2				
4	MASLAN	3	1	3	1	2				
5	MULFAIRI	4	1	4	1	2				
6	MAMARADI K	5	1	4	1	2				
7	AZHARI AGUSTI	6	1	4	1	2				
8	LULIANO	7	1	4	1	2				
9	GESTIANI BULOLO	8	1	4	1	2				
10	SARDIWI	9	1	4	1	2				
11	YOLA	11	1	5	2	2				
13	MERI	11	1	5	2	2				
14	RIMA	12	1	5	2	2				
15	MULAN	12	1	5	2	2				
16	PAUZI	13	1	5	1	2				
17	RINA SALANGAN	14	1	5	1	2				
18	MESI KOSWITA	15	1	7	1	2				
19	PERDIANTO	16	2	7	2	1				
20	MASASRI	17	2	7	2	1				
21	VIARA Z	18	2	7	2	1				
22	RETHA	19	2	7	2	1				
23	DIKI	19	2	7	2	1				
24	WANDI	20	2	7	2	1				
25	DODI	20	2	7	2	1				
26	CHONG	21	2	7	2	1				
27	RICHI	21	2	7	2	1				
28	RUDDI	21	2	7	2	1				
29	RAHMAT	21	2	7	2	1				
30	DEFS	21	2	7	2	1				
31	RUDI	21	2	7	2	1				
32	ADRIZON	13	1	6	1	2				
33	SLAMET	3	1	6	1	2				
34	PUPTRA	21	2	7	2	1				
35	BADI	22	2	7	2	1				
36	PUPTRA	22	2	7	2	1				
37	HARIS	22	2	7	2	1				
38	ERI	23	2	7	3	1				
39	HADI	23	2	7	3	1				
40	HENDRA	23	2	7	3	1				
41	HOP	23	2	7	3	1				
42	JAMRI	20	2	7	2	1				
43	CHAN	20	2	7	2	1				
44	ARIP	24	2	7	2	1				
45	DONI	24	2	7	2	1				
46	JOHN	25	2	7	2	1				
47	ANAN	25	2	7	2	1				
48	SHAFRIZAL	26	2	7	2	1				
49	AZHAR	26	2	7	2	1				
50	VINA	26	2	7	2	1				
51	DINA	27	2	5	4	2				
52	SIKA	27	2	5	4	2				
53	MAYAR	28	2	5	1	1				
54	LASTRI	28	2	7	1	1				
55	MERY	28	2	7	1	1				
56	ADRIZON	28	2	7	1	1				
57	MULENDRI ADI	13	1	6	1	2				
58	ADRIZON	13	1	6	1	2				
59	MASASRI	29	2	7	2	1				
60	DIKI	29	2	7	2	1				

Setelah dilakukan iterasi ke-1 maka lakukan iterasi ke-2 untuk membandingkan letak posisi hasil cluster 1 dan cluster 2. Jika posisinya tidak berubah maka iterasi berikutnya tidak perlu dilakukan lagi. Untuk melakukan iterasi ke-2 maka tentukan cluster baru terlebih dahulu.

4. Melakukan Pencarian Iterasi 2

Pada pembahasan ini dilakukan pencarian iterasi 2, dengan melakukan perhitungan sebagai berikut

Tabel 2.3 Pusat Cluster Iterasi 2

Pusatcluster 1	23,361	2	6,833	2,111	1,056
Pusatcluster 2	9,25	1,083	4,667	1,25	1,958

a. Perhitungan jarak data 1 terhadap pusat cluster.

Perhitungan pusat cluster terhadap data 1 dengan penjelasan sebagai berikut:

$$D(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ik} - X_{jk})^2}$$

Perhitungan gaji dari data ke-1 terhadap pusat cluster :

$$\text{Centroid 1} = \frac{\sqrt{(1-23,361)^2 + (1-2)^2 + (1-6,833)^2 + (1-2,111)^2 + (2-1,056)^2}}{537.174}$$

$$\text{Centroid 2} = \frac{\sqrt{(2-23,361)^2 + (1-2)^2 + (2-6,833)^2 + (1-2,111)^2 + (3-1,056)^2}}{485.674}$$

$$\text{Centroid 3} = \frac{\sqrt{(3-23,361)^2 + (1-2)^2 + (3-6,833)^2 + (1-2,111)^2 + (2-1,056)^2}}{432.396}$$

$$\text{Centroid 4} = \frac{\sqrt{(3-23,361)^2 + (1-2)^2 + (3-6,833)^2 + (1-2,111)^2 + (2-1,056)^2}}{432.396}$$

$$\text{Centroid 5} = \frac{\sqrt{(4-23,361)^2 + (1-2)^2 + (4-6,833)^2 + (1-2,111)^2 + (2-1,056)^2}}{386.007}$$

Perhitungan gaji dari data ke-2 terhadap pusat cluster :

$$\text{Centroid 1} = \frac{\sqrt{(1-9,25)^2 + (1-1,083)^2 + (1-4,667)^2 + (1-1,125)^2 + (2-1,958)^2}}{81.578}$$

$$\text{Centroid 2} = \frac{\sqrt{(2-9,25)^2 + (1-1,083)^2 + (2-4,667)^2 + (1-1,125)^2 + (2-1,958)^2}}{60.828}$$

$$\text{Centroid 3} = \frac{\sqrt{(3-9,25)^2 + (1-1,083)^2 + (3-4,667)^2 + (1-1,125)^2 + (2-1,958)^2}}{41.911}$$

$$\text{Centroid 4} = \frac{\sqrt{(3-9,25)^2 + (1-1,083)^2 + (3-4,667)^2 + (1-1,125)^2 + (2-1,958)^2}}{41.911}$$

$$\text{Centroid 5} = \frac{\sqrt{(4-9,25)^2 + (1-1,083)^2 + (4-4,667)^2 + (1-1,125)^2 + (2-1,958)^2}}{28.078}$$

Hingga seterusnya data diperhitungkan sampai data yang 60 dan hasil dari data perhitungan yang dilakukan sebagai berikut.

Tabel 4.5 : Hasil Perhitungan Gaji dan Pengelompokan Data Iterasi ke-2

No	NAMA	Bagian	STATUS	Gaji Pokok	JUMLAH (Yang Diterima)	centroid 1				
						23.361	2	6.833	2.111	1.056
1	RUDI DARMAWI	1	1	1	1	2				
2	RINDI MAHER	2	1	2	1	3				
3	MAMARADI	3	1	3	1	2				
4	MASLAN	3	1	3	1	2				
5	MULFAIRI	4	1	4	1	2				
6	MAMARADI K	5	1	4	1	2				
7	AZHARI AGUSTI	6	1	4	1	2				
8	LULIANO	7	1	4	1	2				
9	GESTIANI BULOLO	8	1	4	1	2				
10	SARDIWI	9	1	4	1	2				
11	YOLA	11	1	5	2	2				
13	MERI	11	1	5	2	2				
14	RIMA	12	1	5	2	2				
15	MULAN	12	1	5	2	2				
16	PAUZI	13	1	5	1	2				
17	RINA SALANGAN	14	1	5	1	2				
18	MESI KOSWITA	15	1	7	1	2				
19	PERDIANTO	16	2	7	2	1				
20	MASASRI	17	2	7	2	1				
21	VIARA Z	18	2	7	2	1				
22	RETHA	19	2	7	2	1				
23	DIKI	19	2	7	2	1				
24	WANDI	20	2	7	2	1				
25	DODI	20	2	7	2	1				
26	CHONG	21	2	7	2	1				
27	RICHI	21	2	7	2	1				
28	RUDDI	21	2	7	2	1				
29	RAHMAT	21	2	7	2	1				
30	DEFS	21	2	7	2	1				

Setelah dilakukan iterasi ke-3 ternyata hasilnya masih berbeda, maka lakukan iterasi ke-4 untuk membandingkan letak posisi hasil cluster 1 dan cluster 2. Jika posisinya tidak berubah maka iterasi berikutnya tidak perlu dilakukan lagi. Untuk melakukan iterasi ke-4 maka tentukan cluster baru terlebih dahulu.

Didapatkan hasil *cluster* baru sebagai berikut:

centroid 2				
8.591	1	4.455	1.182	2.045
centroid 2				
23	2	6.842	2.105	1.053

6. Melakukan Pencarian Interasi 4.

Pada pembahasan ini dilakukan pencarian interasi 4, dengan melakukan perhitungan sebagai berikut : Perhitungan gaji data 1 terhadap pusat cluster.

Pusat cluster 1	8,591	1,000	4,455	1,182	2,045
Pusat cluster 2	23,000	2	6,842	2,105	1,053

Perhitungan pusat *cluster* terhadap data 1 dengan penjelasan sebagai berikut:

$$Dij = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Perhitungan gaji dari data ke-1 terhadap pusat *cluster* :

centroid 1 = $\frac{\sqrt{(1 - 8,591)^2 + (1 - 1)^2 + (1 - 4,455)^2 + (1 - 1,182)^2 + (2 - 2,045)^2}}{69,591}$

Centroid 2 = $\frac{\sqrt{(2 - 8,591)^2 + (1 - 1)^2 + (2 - 4,455)^2 + (1 - 1,182)^2 + (3 - 2,045)^2}}{50,409}$

Centroid 3 = $\frac{\sqrt{(3 - 8,591)^2 + (1 - 1)^2 + (3 - 4,455)^2 + (1 - 1,182)^2 + (2 - 2,045)^2}}{33,409}$

Centroid 4 = $\frac{\sqrt{(3 - 8,591)^2 + (1 - 1)^2 + (3 - 4,455)^2 + (1 - 1,182)^2 + (2 - 2,045)^2}}{33,409}$

Centroid 5 = $\frac{\sqrt{(4 - 8,591)^2 + (1 - 1)^2 + (4 - 4,455)^2 + (1 - 1,182)^2 + (2 - 2,045)^2}}{21,318}$

Perhitungan gaji dari data ke-2 terhadap pusat *cluster* :

Centroid 1 = $\sqrt{(1 - 23)^2 + (1 - 2)^2 + (1 - 6,842)^2 + (1 - 2,105)^2 + (2 - 1,053)^2} = 521,249$

Centroid 2 = $\sqrt{(2 - 23)^2 + (1 - 2)^2 + (2 - 6,842)^2 + (1 - 2,105)^2 + (3 - 1,053)^2} = 470,460$

Centroid 3 = $\sqrt{(3 - 23)^2 + (1 - 2)^2 + (3 - 6,842)^2 + (1 - 2,105)^2 + (2 - 1,053)^2} = 417,881$

Centroid 4 = $\sqrt{(3 - 23)^2 + (1 - 2)^2 + (3 - 6,842)^2 + (1 - 2,105)^2 + (2 - 1,053)^2} = 417,881$

Centroid 5 = $\sqrt{(4 - 23)^2 + (1 - 2)^2 + (4 - 6,842)^2 + (1 - 2,105)^2 + (2 - 1,053)^2} = 372,197$

Hingga seterusnya data diperhitungkan sampai data yang 60, Setelah didapatkan *cluster* baru pada iterasi ke-3 maka dilakukan *clustering* pada interasin ke-4 yang mana hasil *cluster* tidak mengalami perubahan terlihat hasil perhitungan pada table 4.6 sehingga perulangan dihentikan. Dan Hasil dari data perhitungan yang dilakukan sebagai berikut

Tabel 4.7. :Hasil Perhitungan Gaji dan Pengelompokan Data Interasi ke-4

No	Bagian	STATUS	Gaji Pokok	Jumlah Hari Kerja	Uraian
1	23010	1	8.591	1	1
2	23010	1	8.591	1	1
3	23010	1	8.591	1	1
4	23010	1	8.591	1	1
5	23010	1	8.591	1	1
6	23010	1	8.591	1	1
7	23010	1	8.591	1	1
8	23010	1	8.591	1	1
9	23010	1	8.591	1	1
10	23010	1	8.591	1	1
11	23010	1	8.591	1	1
12	23010	1	8.591	1	1
13	23010	1	8.591	1	1
14	23010	1	8.591	1	1
15	23010	1	8.591	1	1
16	23010	1	8.591	1	1
17	23010	1	8.591	1	1
18	23010	1	8.591	1	1
19	23010	1	8.591	1	1
20	23010	1	8.591	1	1
21	23010	1	8.591	1	1
22	23010	1	8.591	1	1
23	23010	2	6.842	2	2
24	23010	2	6.842	2	2
25	23010	2	6.842	2	2
26	23010	2	6.842	2	2
27	23010	2	6.842	2	2
28	23010	2	6.842	2	2
29	23010	2	6.842	2	2
30	23010	2	6.842	2	2
31	23010	2	6.842	2	2
32	23010	2	6.842	2	2
33	23010	2	6.842	2	2
34	23010	2	6.842	2	2
35	23010	2	6.842	2	2
36	23010	2	6.842	2	2
37	23010	2	6.842	2	2
38	23010	2	6.842	2	2
39	23010	2	6.842	2	2
40	23010	2	6.842	2	2
41	23010	2	6.842	2	2
42	23010	2	6.842	2	2
43	23010	2	6.842	2	2
44	23010	2	6.842	2	2
45	23010	2	6.842	2	2
46	23010	2	6.842	2	2
47	23010	2	6.842	2	2
48	23010	2	6.842	2	2
49	23010	2	6.842	2	2
50	23010	2	6.842	2	2
51	23010	2	6.842	2	2
52	23010	2	6.842	2	2
53	23010	2	6.842	2	2
54	23010	2	6.842	2	2
55	23010	2	6.842	2	2
56	23010	2	6.842	2	2
57	23010	2	6.842	2	2
58	23010	2	6.842	2	2
59	23010	2	6.842	2	2
60	23010	2	6.842	2	2

Pada iterasi ke-3 dan iterasi ke-4 hasil cluster tidak berubah maka perulangan tidak dilakukan lagi. Dapat ditarik kesimpulan bahwa cluster 0 terdiri 22 orang karyawan, cluster 1 terdiri dari 38 orang karyawan . Terlihat pada Tabel berikut 4.8:

4.8 : Hasil Akhir Cluster

	NAMA	BAGIAN	STATUS	GAJI POKOK	Jumlah Hari Kerja	TOTAL
Cluster	RUDI DARMAWI	1	1	1	1	2
	RIKO MAYER	2	1	2	1	3
	MAWARDI	3	1	3	1	2
	MASLAN	3	1	3	1	2
	MULFAJRI	4	1	4	1	2
	MAWARDI K	5	1	4	1	2
	AZHARI AGUS TI	6	1	4	1	2
	JULIANO	7	1	4	1	2
	GESTIANI BULOLO	8	1	4	1	2
	SARDEWI	9	1	4	1	2
	ANDHIKA	10	1	4	1	2
	YOLA	11	1	5	2	2
	NIERI	12	1	5	2	2
RIMA	12	1	5	2	2	
WULAN	12	1	5	2	2	
FAUZI	13	1	5	1	2	



NAMA	BAGIAN	STATUS	GAJI POKOK	JUMLAH HARI KERJA	TOTAL
GAJI KARYAWAN TINGGI					
RUDI DARMAWI	1	1	1	1	2
RIKO MAYER	2	1	2	1	3
MAWARDI	3	1	3	1	2
MASLAN	3	1	3	1	2
MULFAJRI	4	1	4	1	2
MAWARDI K	5	1	4	1	2
AZHARI AGUSTI	6	1	4	1	2
JULIANO	7	1	4	1	2
GESTIANI BUULOLO	8	1	4	1	2
SARDEWI	9	1	4	1	2
ANDHIKA	10	1	4	1	2
YOLA	11	1	5	2	2
MERI	12	1	5	2	2
RIMA	12	1	5	2	2
WULAN	12	1	5	2	2
FAUZI	13	1	5	1	2
GAJI KARYAWAN RENDAH					
RINA SIALANGAN	14	1	5	1	2
MESI KOSWITA	15	1	7	1	2
ADRIZON	13	1	6	1	2
SLAMET	3	1	6	1	2
TULENDRI ADI	13	1	6	1	2
ADRIZON	13	1	6	1	2
FERDIANTO	16	2	7	2	1
MAISARIL	17	2	7	2	1
VIARA Z	18	2	7	2	1
RETHA	19	2	7	2	1
OKY	19	2	7	2	1
WANDI	20	2	7	2	1
DODI	20	2	7	2	1
OYONG	21	2	7	2	1
RICHI	21	2	7	2	1
RUDDI	21	2	7	2	1
RAHMAT	21	2	7	2	1
DEFIS	21	2	7	2	1
RUDI	21	2	7	2	1
PUTRA	21	2	7	2	1
BADI	22	2	7	2	1
VITRA	22	2	7	2	1
HARIS	22	2	7	2	1
ERI	23	2	7	3	1
HADI	23	2	7	3	1
HENDRA	23	2	7	3	1
YOPI	23	2	7	3	1
JAMRI	20	2	7	2	1
MERI	28	2	7	1	1
ADRIZON	28	2	7	1	1
MAISARIL	29	2	7	2	1
OKY	29	2	7	2	1

V. IMPLEMENTASI DAN PENGUJIAN

Setelah dilakukan penganalisaan data seperti penjelasan pada bab sebelumnya maka langkah selanjutnya adalah melakukan implementasi. Implementasi dilakukan dengan menggunakan aplikasi *RapidMiner* 6.2 yang akan membantu mempermudah dalam pengelompokan gaji karyawan tetap dan kontrak yang akan membantu dalam proses pengambilan keputusan.

A. Hasil Pengujian Clustering

Pengujian terhadap analisa sangat penting di lakukan untuk menemukan dan memastikan apakah hasil dari analisa secara manual menggunakan *data mining* algoritma *K-Means* sesuai dengan keputusan yang di harapkan dengan pengujian data menggunakan *software Rapid Miner 7.5*.

Pada Cluster Model (Clustering) dapat dilihat jumlah dari ke-4cluster. Pada Cluster Model (Clustering) ini dapat dilihat jumlah dari ke-2cluster adalah cluster 1 sebanyak 22items.

Cluster2sebanyak 38items. Keseluruhan data berjumlah 60 items

Cluster Model

Cluster 0: 22 items
Cluster 1: 38 items
Total number of items: 60

Gambar 3 Tampilan Cluster Model (Text View)

Pada centroid table dapat kita lihat bahwa cluster1 merupakan kelompok karyawan yang bergaji rendah, dan cluster2 kelompok karyawan yang bergaji tinggi. Hal ini dapat dibaca pada centroid masing-masing cluster pada attribute nilai rata-rata.

Attribute	cluster_0	cluster_1
Bagian	8.591	23
STATUS	1.136	1.658
Gaji Pokok	4.500	6.289
JUMLAH HAI	1.318	1.868
Yang Diterim	1.955	1.447

Gambar 4 Tampilan Centroid Table

VI. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan analisa dan pengujian menggunakan *Software RapidMiner* pada PT Indomex Dwijaya Lestari, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Metode Clustering K-Means dapat membantu mengelompokkan gaji karyawan berdasarkan faktor-faktor yang mempengaruhi kinerja karyawan dalam bekerja sehingga membantu proses administrasi PT Indomex Dwijaya Lestari.
2. Didapatkan 2 kategori gaji karyawan, yaitu bergaji kategori tinggi dan bergaji kategori rendah dimana masing-masing cluster memiliki nilai rata-rata yang berbeda.
3. Penggunaan Algoritma k-means Clustering dapat membantu dan mempermudah dalam pencarian informasi
4. Aplikasi RapidMiner mampu mengatasi persoalan-persoalan dalam proses pengelompokan gaji karyawan, dari pada melakukan pencarian hasil secara manual karena kesalahan dalam proses pencarian oleh Administrator.

B. Saran

Untuk pengembangan dalam penelitian berikutnya maka dituliskan saran-saran sebagai berikut :

1. Faktor-faktor yang mempengaruhi rendah atau tingginya gaji karyawan dapat ditambahkan sehingga pemahaman terhadap data karyawan lebih luas lagi.

2. Pada pengembangan penelitian selanjutnya dapat dilakukan dengan jumlah data yang lebih besar lagi.
3. Aplikasi RapidMiner ini merupakan suatu software yang userfriendly dan dapat digunakan untuk membantu mengelompokkan data.
4. Software RipadMiner dimanfaatkan pada bidang yang lebih luas lagi. Karena dengan menggunakan software ini dalam aplikasi pekerjaan sehari-hari dapat membantu aktifitas pekerjaan

Algorithm for Multicore Heterogeneous Compute Cluster. International Journal of Information and Communication Technology Research ISSN 2223-4985

Sankar Rajagopal. 2011. Customer data Clustering using data mining technique. Enterprise DW/BI Consultant Tata Consultancy Services, Newark, DE, USA

REFERENSI

- Ahmad Sanmorino. 2012. Clustering Batik Images using Fuzzy C-Means Algorithm Based on Log-Average Luminance. Faculty of Computer Science, Universitas Indonesia.
- Amita Verma, Ashwani kumar. 2014. Performance Enhancement of K-Means Clustering Algorithms for High Dimensional Data sets. *International Journal of Advanced Research in Computer Science and Software Engineering Research Paper* Available online at: www.ijarcsse.com
- Angga Ginanjar Mabur, Riani Lubis. 2012. Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia.
- Emerensye S. Y. Pandie. 2012. Implementasi Algoritma data mining K-Nearest Neighbor (K-NN) dalam pengambilan keputusan pengajuan kredit. *Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana*
- J Oyelade O, O Oladipupo O, C Obagbuwa I (2010), "Penerapan algoritma K-Means Clustering untuk prediksi kinerja akademik mahasiswa"
- Noticewala Maitry dan Dinesh Vaghela. 2014. Survey on Different Density Based Algorithms on Spatial Dataset. *International Journal of Advance Research in Computer Science and Management Studies Research Article / Paper / Case Study* Available online at: www.ijarcsms.com
- Nurhayati. 2014. Metode Rough Set untuk melihat perilaku suami yang menjadi akseptor KB vasetomi. Program Studi Teknik Informatika, STMIK Kaputama Binjai.
- Ramesh Singh Yadava1, P.K.Mishra2. 2012. Performance Analysis of High Performance k-Mean Data Mining