

DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASTERISASI TINGKAT TRIDARMA PENGAJARAN DOSEN

Rizki Muliono¹, Zulfikar Sembiring²

¹² Program Studi Teknik Informatika, Universitas Medan Area
Jl. Kolam No.1, Medan Estate, Sumatera Utara

¹rizkimuliono@gmail.com, ²zoelsembiring@gmail.com

Page | 272

Abstrak— Universitas medan area memiliki dosen dengan jumlah yang banyak dimana setiap dosen mengampuh matakuliah sesuai bidang keahliannya masing-masing. Setiap semesternya dosen diwajibkan membuat dokumen pengajaran seperti Silabus, RPS, Kontrak Kuliah, RPP dan kemudian diupload ke aplikasi RPS milik Universitas Medan Area untuk dinilai oleh Unit LP2MP yang memiliki tugas malakukan klasterisasi terhadap hasil pembobotan nilai dari tiap-tiap dokumen dari dosen-dosen. Hasil klasterisasi tersebut selanjutnya akan merujuk pada pemberian besaran nilai tunjangan yang di berikan kepada dosen yang membuat dan mengumpulkan dokumen-dokumen pengajaran tesebut. Untuk membantu klasterisasi digunakan algoritma *K-Means Clustering* adalah salah satu teknik dari data mining dengan metode clustering non hirarki didalam prosesnya berusaha mempartisi data-data yang ada ke dalam bentuk klaster. Penelitian ini diharapkan dapat membantu proses klasterisasi dengan nilai yang mendekati karakteristik menjadi lebih efektif. Ketepatan prediksi yang dilakukan oleh algortima *K-means* terhadap 15 data mengalami perbedaan ketepatan, hanya sebanyak 53.33% akurasi prediksi bernilai benar.

Keywords— Data Mining, Algoritma *K-Means Clustering*, Tingkat Tridarma Dosen, Klasterisasi

Abstract— The field area university has a large number of lecturers in which each lecturer applies the subject according to their respective fields of expertise. Every semester the lecturers are required to make teaching documents such as Syllabus, RPS, Lecture Contracts, RPP and then upload to the RPS application owned by the University of Medan Area to be assessed by the LP2MP Unit who has the task of clustering the results of weighting each document from the lecturers. The results of the clustering will then refer to the provision of the amount of the allowance given to the lecturer who made and collected the teaching documents. To help with clustering, the *K-Means Clustering* algorithm is used as one of the techniques of data mining with non-hierarchical clustering methods in the process of trying to partition existing data into clusters. This research is expected to help the clustering process with values that are close to the characteristics to be more effective. The accuracy of predictions made by the *K-means* algorithm for 15 data experiences different accuracy, only as much as 53.33% predictive accuracy is true.

Keywords— Data Mining, *K-Means Clustering Algorithm*, Level Tridarma Lecturer, Clustering

I. PENDAHULUAN

Tridarma adalah tugas utama bagi orang yang berprofesi sebagai dosen, tugas dari tridarma tersebut meliputi pengajaran, penelitian dan pengabdian kepada masyarakat. Kinerja seorang dosen dinilai dari poin-poin yang dihasilkan dari parameter yang ada didalam tugas-tugas tidap tridarma tersebut. Sehingga keaktifan dosen dapat dilihat dari nilai tridarma perguruan tinggi yang telah dia lakukan. Indikator inilah yang menjadikan Universitas Medan Area membuat kebijakan wajib bagi setiap dosen-dosen nya untuk melakukan tridarma perguruan tinggi. Pada pengajaran memiliki beberapa point yang menjadi indikator penilaian beberapa diantaranya yang diwajibkan adalah dokumen Silabus matakuliah, Rencana Pengajaran

Semesteran (RPS), Kontrak Kuliah (KK), Rencana Proses Pembelajaran (RPP) dan kemudian diupload ke sistem informasi RPS.

Untuk menjamin mutu materi dan proses pengajaran yang terjadi di Universitas Medan Area, semua dokumen-dokumen tersebut akan dinilai dan dikoreksi oleh unit khusus bidang penjaminan mutu akademik yaitu Unit LP2MP dimana salah satu tugasnya adalah melakukan koreksi dan penilaian terhadap dokumen-dokumen tersebut, yang sebelumnya di himpun melalui sebuah sistem informasi RPS secara online. Dosen yang melengkapi dokumen-dokumen tersebut dan upload ke sistem informasi RPS akan diberikan reward berupa tunjangan tambahan penghasilan yang nominalnya

disesuaikan dengan nilai hasil klusterisasi yang dilakukan oleh Unit LP2MP tersebut.

K-Means adalah salah satu algoritma dari teknik data mining yang mampu melakukan klusterisasi terhadap data heterogen karena pada dasarnya algoritma pengelompokan hanya mampu mengenali nilai atribut homogen saja [1]. Proses algoritma *k-means* berbeda dengan dengan algoritma data mining lainnya seperti Algoritma Apriori yang mencari *frequent itemsets* yang sering muncul dengan model data heterogen juga kemudian di pangkas dan di lakukan perhitungan sesuai jumlah *K* itemsetnya [2], klusterisasi mengelompokkan sejumlah *n* buah objek kedalam *k* kelas dengan berdasarkan perhitungan jarak nya dengan pusat cluster [3]. Dalam beberapa penelitian sebelumnya telah berhasil melakukan klusterisasi data dengan algoritma *k-means* untuk kasus prediksi waktu kelulusan mahasiswa berdasarkan parameter IPK dan kehadiran, yang selanjutnya menggunakan teknik Analisis cluster yaitu mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya, mengklasifikasi titik objek sehingga setiap objek yang paling dekat kesamaannya dengan titik objek lain yang berada dalam cluster yang sama [4].

Algoritma *K-Means* juga telah berhasil menyelesaikan masalah klusterisasi kinerja akademik mahasiswa yang datanya digali dari tumpukkan data-data seleksi penerimaan mahasiswa baru, dengan hasil proses mining dan teknik clusterisasi yang berjumlah 3 dengan nilai *silhouette coefficient* yang paling mendekati nilai = 1, yaitu nilai 0,108690751. Kemudian pendapatan orang tua tidak mempengaruhi tingkat kinerja akademik dari mahasiswa dan nilai akademis mahasiswa yang telah masuk melalui jalur reguler dan jalur prestasi akademik mempunyai nilai IPK rata-rata tertinggi [5]. Dalam artikel jurnal di [6] menjelaskan kasus yang lain dengan *k-means* untuk *Clustering* untuk Menentukan Nilai Gizi Balita dengan kategori obesitas, gizi lebih, baik, kurang dan gizi buruk pada Posyandu dengan acuan variabel parameter tinggi badan balita dan berat badan, dari hasil penelitiannya ditemukan masih terdapat 30% balita menderita obesitas dan 11% mengalami kekurangan gizi, maka dari itu perlu adanya penanganan dan perhatian khusus dari kader Posyandu dan Puskesmas dilokasi tersebut.

Pada penelitian ini akan di fokuskan pada klusterisasi untuk memprediksi tingkat tridarma seorang dosen melalui penilaian terhadap koreksi dokumen-dokumen yang diupload oleh dosen tersebut, dengan parameter penilaian Silabus, Kontrak Kuliah, RPS, RPP dengan urutan kategori klaster A,B,C,D,E,F. Hasil klusterisasi selanjutnya akan di bandingkan dengan menggunakan data real yang ada untuk dilanjutkan pada analisa *Clustering* terhadap persentase tingkat ketepatan prediksi algoritma *K-Means* dalam

penyelesaian klusterisasi tridarma dosen di Universitas Medan Area dan diharapkan penelitian memiliki manfaat lebih kepada LP2MP sebagai unit pelaksana koreksi dan pengklasteran kinerja tridarma dosen sehingga mutu pendidikan dan pengajaran tetap terjaga.

II. METODOLOGI

Pada penelitian ini dilakukan di lingkungan Universitas Medan Area untuk mengklusterisasi tingkatan kinerja tridarma pembelajaran terhadap dosen. Data yang digunakan adalah data hasil penilaian dokumen-dokumen pendukung pembelajaran dengan format tabulasi data excel. Teknik pengolahan data yang digunakan dalam penelitian ini adalah *Data Mining* dengan metode *K-Means Clustering*. *Tools* yang akan digunakan dalam pengolahan data adalah menggunakan rancangan sistem berbasis web dengan bahasa pemrograman PHP dan selanjutnya hasil prediksi dengan algoritma *K-Means* akan di bandingkan dengan hasil perhitungan data real ada untuk melihat persentase akurasi dari algoritma terhadap data yang di olah.

A. Data Mining

Proses pencarian dan penggalian informasi dari tumpukan-tumpukan data dengan jumlah yang besar merupakan proses utama dari *data mining*, tujuan utama dari pengolahan data tersebut adalah menghasilkan informasi yang baru [2]. Data mining juga disebut dengan *Knowledge Discovery in Database (KDD)* yaitu sebuah proses yang secara otomatis saat pencarian data didalam *space memory* yang amat besar dari data untuk mengetahui pola dengan menggunakan teknik seperti klasifikasi hubungan (*association*) atau pengelompokan (*clustering*) [7]. Oleh sebab itu data mining dapat digunakan untuk mengevaluasi kinerja tridarma dosen dengan menggunakan algoritma yang ada dalam *data mining*, dicoba untuk mengekstrak pengetahuan yang bisa menggambarkan kinerja tridarma dosen pada tiap semester nya.

B. Clustering

Clustering pada suatu data adalah suatu tahapan untuk menggolongkan himpunan data yang atribut kelasnya belum dideskripsikan, secara konsep *clustering* adalah untuk memaksimalkan dan meminimalkan kemiripan intra antar kelas. sebagai contoh, ada suatu himpunan obyek, proses pertama dapat di klusterisasi menjadi beberapa himpunan kelas selanjutnya menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan kelompok klasifikasi tertentu. Cluster juga dapat diartikan sebagai kelompok. Maka analisa clustering pada dasarnya akan menghasilkan sejumlah cluster (kelompok). Sebelum

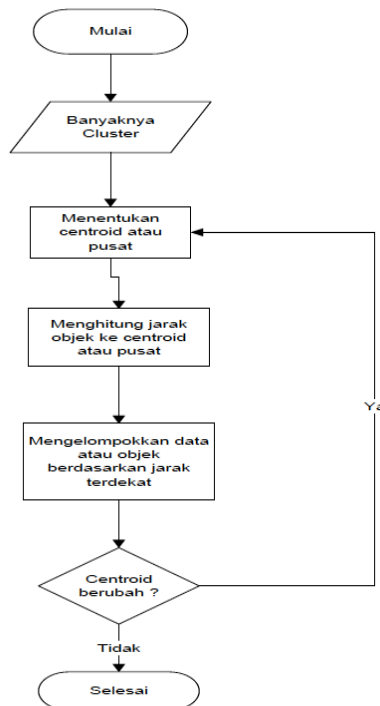
kita melakukan analisa perlu diterapkan pemahaman bahwa suatu himpunan dari data tertentu sebenarnya sudah memiliki kemiripan di antara anggotanya tersebut. Oleh karena itu, setiap anggota-anggota yang memiliki kemiripan karakteristik dikelompokkan ke dalam satu atau lebih dari suatu kelompok [8]. Tujuan dari pada clustering data adalah untuk meminimalisasikan fungsi tujuan yang ditetapkan didalam proses clustering, dan umumnya selalu meminimalisasi variasi suatu cluster dan memaksimalkan variasi antar cluster [9].

C. Algoritma K-Means Clustering

K-Means adalah salah satu teknik clustering pada Data Mining proses pemodelan tanpa supervisi dan metode pengelompokan data secara partisi. Data yang dikelompokkan metode *K-Means* menjadi beberapa kelompok dan setiap kelompok memiliki karakteristik yang mirip atau sama dengan lainnya tetapi dengan kelompok lainnya memiliki karakteristik berbeda. Dengan tujuan meminimalisasi perbedaan setiap data didalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain [10].

Istilah-istilah didalam algoritma k-means clustering :

1. *Cluster* : *Cluster* adalah kelompok atau grup.
2. *Centroid* : *Centroid* adalah titik pusat untuk menentukan *euclidian distance*.
3. Iterasi : Iterasi adalah pengulangan proses, berhenti ketika hasil iterasi telah konvergen.



Gbr. 1 Flowchart proses algoritma K-Means

Langkah-langkah pada algoritma k-means clustering secara umum [4]:

- a. Menentukan jumlah k sebagai cluster yang akan dibentuk. Penentuan banyaknya jumlah cluster k biasanya dilakukan dengan beberapa faktor pertimbangan baik teoritis dan konseptual yang kemudian diusulkan untuk menentukan berapa banyak jumlah cluster.
- b. Bangkitkan k Centroid awal (titik pusat cluster) secara random. Untuk menentukan centroid awal dilakukan dalam bentuk acak dari beberapa objek yang tersedia sebanyak jumlah dari k cluster kemudian untuk menghitung centroid cluster ke-i berikutnya, menggunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n}; \quad i = 1, 2, 3, \dots, n$$

Keterangan :

v = Centroid pada cluster

x_i = Objek ke- i

n = Banyaknya jumlah objek yang menjadi anggota cluster.

- c. Menghitung jarak dari setiap objek ke masing-masing centroid dari masing-masing data cluster dan menghitung jarak antara objek dengan centroid yang bisa dilakukan dengan menggunakan rumus matematik *euclidian distance* :

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; \quad i = 1, 2, 3, \dots, n$$

Keterangan :

x_i = objek x ke- i

y = daya y ke- i

n = banyaknya objek

- d. Hitung masing-masing objek ke dalam centroid yang paling terdekat.
- e. Lakukan iterasi ke- i , kemudian tentukan posisi centroid baru dengan menggunakan persamaan sebelumnya
- f. Ulangi langkah pada poin c jika posisi centroid baru tidak sama dan iterasi atau perulangan akan berhenti jika rasio tidak lebih besar dari nilai rasio sebelumnya hingga hasil perngitungan di masing-masing data *konvergen*.

D. Analisa Data dan Kebutuhan Data

Data yang digunakan pada penelitian ini adalah tabulasi data hasil penilaian dokumen-dokumen pendukung pengajaran matakuliah pada semester ganjil T.A 2018/2019.

TABEL I
DATA SAMPEL HASIL PENILAIAN DOKUMEN

No	Nama Dosen	KK	SL	RPS	RPP	Kat
1	Dosen 1	4.00	9.00	36.00	26.75	B
2	Dosen 2	3.00	4.00	20.00	23.50	E
3	Dosen 3	4.00	9.00	36.00	19.25	C
4	Dosen 4	5.00	6.00	35.00	39.00	A
5	Dosen 5	3.00	6.00	35.00	25.50	C
6	Dosen 6	5.00	9.00	39.00	35.75	A
7	Dosen 7	5.00	8.00	37.00	32.75	B
8	Dosen 8	3.00	4.00	24.00	24.75	D
9	Dosen 9	3.00	4.00	20.00	23.50	E
10	Dosen 10	3.00	10.00	43.00	34.50	A
11	Dosen 11	4.00	9.00	36.00	26.75	B
12	Dosen 12	5.00	10.00	42.50	36.50	A
13	Dosen 13	4.00	6.00	36.00	29.25	B
14	Dosen 14	5.00	8.00	39.00	31.25	B
15	Dosen 15	5.00	8.00	41.00	31.25	A

Keterangan :

KK = Nilai Dokumen Kontrak Kuliah

SL = Nilai Dokumen Silabus

RPS = Nilai Dokumen Rencana Pembelajaran Semesteran

RPP = Nilai Dokumen Rencana Proses Pengajaran

Kat = Kategori Tingkat Penilaian Akhir Dokumen

Adapun range kriteria penilaian yang menjadi peraturan dan acuan bagi Unit pemeriks dan penilai dokumen-dokumen tersebut mengacu pada Panduan Kurikulum Pendidikan Tinggi Universitas Medan Area [11] dapat dilihat pada tabel berikut.

TABEL II
KATEGORI PENILAIAN DOKUMEN

Kategori	Score Penilaian	Σ Nilai Tunjangan
A	100 – 85	500.000

B	75 – 84,9	400.000
C	65 – 74,9	300.000
D	55 – 64,9	200.000
E	45 – 54,9	150.000

Berdasarkan tabel kategori penilaian diatas maka pada penelitian ini akan melakukan klasterisasi sebanyak 5 Cluster.

E. Proses Perhitungan K-means Clustering

Proses ini merupakan simulasi clustering yang diselesaikan dengan metode matematika, yang dilakukan setiap literasi pada sampel data yang ada pada tabel 1.

1. Menentukan *centroid* awal, penentuan ini dilakukan secara acak terhadap tabel data yang ada, untuk titik pusat awal adalah data ke-1 cluster 1, data ke-3 sebagai cluster 2, , data ke-8 cluster 3, data ke-9 sebagai cluster 4 dan data ke-12 sebagai cluster ke 5.

TABEL III
NILAI CENTROID AWAL

No	Nama Dosen	KK	SL	RPS	RPP	Kat
1	Dosen 1	4.00	9.00	36.00	26.75	B
3	Dosen 3	4.00	9.00	36.00	19.25	C
8	Dosen 8	3.00	4.00	24.00	24.75	D
9	Dosen 9	3.00	4.00	20.00	23.50	E
12	Dosen 12	5.00	10.00	42.50	36.50	A

2. Selanjutnya melakukan perhitungan jarak titik pusat *cluster* dengan menggunakan *Euclidian Distance* untuk mendapatkan masing-masing jarak antara kluster C1,C2,C3,C4,C5 dan C6. Data ke-1 dihitung dengan data pusat cluster 1 dengan hasil 0, Data ke-1 dihitung dengan data pusat cluster 1 dengan hasil 0, Data ke-2 dihitung dengan data pusat cluster 1 dengan hasil 17.10, Data ke-3 dihitung dengan data pusat cluster 1 dengan hasil 7.50 dan seterusnya. Rumus perhitungan *euclidian distance* pada *literasi-1* pada Cluster-1 dapat dilihat di tabel IV di bawah ini.

TABEL IV
EUCLIDIAN DISTANCE ITERASI 1 CLUSTER 1 (C1)

C	Perhitungan Euclidian Distance	Σ
1	$\sqrt{(4-4)^2 + (9-9)^2 + (36-36)^2 + (26.75-26.75)^2}$	0.00
1	$\sqrt{(3-4)^2 + (4-9)^2 + (20-36)^2 + (23.5-26.75)^2}$	17.10
1	$\sqrt{(4-4)^2 + (9-9)^2 + (36-36)^2 + (19.25-26.75)^2}$	7.50

1	$\sqrt{(5-4)^2 + (6-9)^2 + (35-36)^2 + (35-26.75)^2}$	12.6 9
1	$\sqrt{(3-4)^2 + (6-9)^2 + (35-36)^2 + (35-26.75)^2}$	3.54
1	$\sqrt{(5-4)^2 + (9-9)^2 + (39-36)^2 + (39-26.75)^2}$	9.54
1	$\sqrt{(5-4)^2 + (8-9)^2 + (37-36)^2 + (37-26.75)^2}$	6.24
1	$\sqrt{(3-4)^2 + (4-9)^2 + (24-36)^2 + (24-26.75)^2}$	13.1 9
1	$\sqrt{(3-4)^2 + (4-9)^2 + (20-36)^2 + (20-26.75)^2}$	17.1 0
1	$\sqrt{(3-4)^2 + (10-9)^2 + (43-36)^2 + (43-26.75)^2}$	10.5 4
1	$\sqrt{(4-4)^2 + (9-9)^2 + (36-36)^2 + (36-26.75)^2}$	0.00
1	$\sqrt{(5-4)^2 + (10-9)^2 + (42.5-36)^2 + (42.5-26.75)^2}$	11.8 0
1	$\sqrt{(4-4)^2 + (6-9)^2 + (36-36)^2 + (36-26.75)^2}$	3.91
1	$\sqrt{(5-4)^2 + (8-9)^2 + (39-36)^2 + (39-26.75)^2}$	5.59
1	$\sqrt{(5-4)^2 + (8-9)^2 + (41-36)^2 + (41-26.75)^2}$	6.87

Hal yang sama dilakukan pada perhitungan jarak antara titik pusat cluster dengan setiap data di hitung dengan titik pusat cluster ke-2 yaitu baris data ke-3 pada Cluster 2 (C2) dilihat pada tabel V dibawah ini.

TABEL V
EUCLIDIAN DISTANCE ITERASI 2 CLUSTER (C2)

C	Perhitungan Euclidian Distance	Σ
2	$\sqrt{(4-4)^2 + (9-9)^2 + (36-36)^2 + (26.75-19.25)^2}$	7.50
2	$\sqrt{(3-4)^2 + (4-9)^2 + (20-36)^2 + (23.5-19.25)^2}$	17.3 2
2	$\sqrt{(4-4)^2 + (9-9)^2 + (36-36)^2 + (19.25-19.25)^2}$	0.00
2	$\sqrt{(5-4)^2 + (6-9)^2 + (35-36)^2 + (39-19.25)^2}$	20.0 3
2	$\sqrt{(3-4)^2 + (6-9)^2 + (35-36)^2 + (25.5-19.25)^2}$	7.08
2	$\sqrt{(5-4)^2 + (9-9)^2 + (39-36)^2 + (35.75-19.25)^2}$	16.8 0
2	$\sqrt{(5-4)^2 + (8-9)^2 + (37-36)^2 + (32.75-19.25)^2}$	13.6 1
2	$\sqrt{(3-4)^2 + (4-9)^2 + (24-36)^2 + (24.75-19.25)^2}$	14.1 5
2	$\sqrt{(3-4)^2 + (4-9)^2 + (20-36)^2 + (23.5-19.25)^2}$	17.3 2
2	$\sqrt{(3-4)^2 + (10-9)^2 + (43-36)^2 + (34.5-19.25)^2}$	16.8 4
2	$\sqrt{(4-4)^2 + (9-9)^2 + (36-36)^2 + (26.75-19.25)^2}$	7.50
2	$\sqrt{(5-4)^2 + (10-9)^2 + (42.5-36)^2 + (36.5-19.25)^2}$	18.4 9
2	$\sqrt{(4-4)^2 + (6-9)^2 + (36-36)^2 + (29.25-19.25)^2}$	10.4 4
2	$\sqrt{(5-4)^2 + (8-9)^2 + (39-36)^2 + (31.25-19.25)^2}$	12.4 5
2	$\sqrt{(5-4)^2 + (8-9)^2 + (41-36)^2 + (31.25-19.25)^2}$	13.0 8

Begitu selanjut dengan hal yang sama untuk perhitungan euclidian distance cluster C3, C4 dan C5 sampai hasil akhir perhitungan semua cluster dapat dilihat pada tabel VI di berikut ini.

TABEL VI
PERHITUNGAN ITERASI 1 SEMUA CLUSTER

ke-	C 1	C 2	C 3	C 4	C 5	Min Jarak	Min Jarak ²
1	0.00	7.50	13.19	17.10	11.80	0.00	0.00
2	17.10	17.32	4.19	0.00	26.74	0.00	0.00
3	7.50	0.00	14.15	17.32	18.49	0.00	0.00
4	12.69	20.03	18.22	21.75	8.86	8.86	78.50
5	3.54	7.08	11.21	15.26	14.04	3.54	12.56
6	9.54	16.80	19.36	23.24	3.72	3.72	13.81
7	6.24	13.61	15.91	19.86	6.95	6.24	39.00
8	13.19	14.15	0.00	4.19	22.81	0.00	0.00
9	17.10	17.32	4.19	0.00	26.74	0.00	0.00
10	10.54	16.84	22.18	26.19	2.87	2.87	8.25
11	0.00	7.50	13.19	17.10	11.80	0.00	0.00
12	11.80	18.49	22.81	26.74	0.00	0.00	0.00
13	3.91	10.44	13.01	17.15	10.57	3.91	15.25
14	5.59	12.45	16.95	21.00	6.62	5.59	31.25
15	6.87	13.08	18.74	22.83	5.81	5.81	33.81

Pada perhitungan iterasi 1 menghasilkan jumlah Σ Jarak Minimum Kuadrat atau WCV = 232.44.

3. Pengelompokkan cluster dari jarak hasil perhitungan dibandingkan dan dipilih jarak yang terdekat antar data dengan pusat cluster nya, sehingga kemudian jarak ini menentukan data tersebut berada dalam satu kelompok dengan pusat cluster yang terdekat dengannya. Hasil pengelompokkan pada literasi 1 bisa dilihat pada tabel dibawah ini.

TABEL VII
PERHITUNGAN ITERASI 1 SEMUA CLUSTER

Dosen	C1	C2	C3	C4	C5
ke 1	X				
ke 2				X	
ke 3		X			
ke 4	X				
ke 5	X				
ke 6	X				
ke 7	X				
ke 8			X		
ke 9				X	
ke 10	X				
ke 11	X				
ke 12	X				

ke 13	X				
ke 14	X				
ke 15	X				

4. Menghitung jarak antar cluster yang dengan kombinasi jarak antara C1 ke C2, C1 ke C3, C1 ke C4, C1 ke C5 dan seterusnya, perhitungan menggunakan rumus *euclidian distance*.

TABEL VIII
PERHITUNGAN ITERASI 1 JARAK ANTAR CLUSTER

Dari	Ke	Jarak
C1	C2	7.50
C1	C3	13.19
C1	C4	17.10
C1	C5	11.80
C2	C3	14.15
C2	C4	17.32
C2	C5	18.49
C3	C4	4.19
C3	C5	22.81
C4	C5	26.74
	Σ BCV	153.31

Dari hasil perhitungan jarak antar cluster di temukanlah jumlah total dari masing-masing jarak antar cluster dengan variabel BCV = 153.000.

5. Melakukan penghitungan nilai rasio yang menjadi acuan untuk melakukan iterasi berikutnya, nilai rasio sekarang dengan nilai rasio sebelumnya, nilai Rasio = WCV/BCV maka $232.44/153.31 = 0.66$. Aturan untuk melakukan literasi berikutnya adalah jika rasio sekarang (0.66) > rasio sebelum (0), maka iterasi ke-2 dilanjutkan.

6. Memperbaharui nilai dari titik pusat cluster baru atau centroid, dengan cara menghitung nilai yang ada grup clustering pada tiap variabel di setiap cluster. Sehingga ditemukan titik cluster baru :

TABEL IX
NILAI CENTROID BARU

Pusat Cluster Baru	KK	SL	RPS	RPP
C1	4.36	8.09	38.14	31.75
C2	4.00	9.00	36.00	19.25
C3	3.00	4.00	36.00	24.75
C4	3.00	3.00	4.50	4.00
C5	0	0	0	0

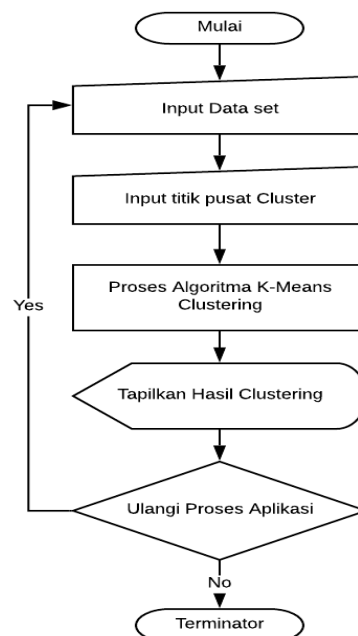
Lanjtkan iterasi ke-2 dengan mengulagi proses perhitungan jarak antar data dengan nilai centroid baru yang ada pada tabel X diatas. Iterasi akan berhenti jika rasio sekarang TIDAK > rasio sebelum nya.

III. PERANCANGAN

Pada penelitian pengolahan data akan dilakukan menggunakan implementasi berbasis web dengan bahasa pemrograman PHP.

1. Prosedur Kerja Sistem

Implementasi sistem yang dirancang di buat dengan sederhana, dapat dilihat pada gambar dibawah ini.



Gbr.2 Prosedur Kerja Sistem yang dirancang

Sistem hanya akan melakukan input data yang akan di clusterisasi kemudian input data yang menjadi nilai titik cluster pusat dan selanjutnya di proses, hasil pengolahan algoritma yang di implementasi dengan pemrograman web PHP disajikan dalam bentuk tabel hitungan *euclidian distance* per-iterasi berikut tampilan pengelompokan data clusteranya.

2. Implementasi Sistem

Pada penelitian ini implementasi input data pada cluster menggunakan data array dengan format seperti di bawah ini :

```

$data = array(
    array(4.00,9.00,36.00,26.75),
    array(3.00,4.00,20.00,23.50),
    array(4.00,9.00,36.00,19.25),
    array(5.00,6.00,35.00,39.00),
    array(3.00,6.00,35.00,25.50),
    array(5.00,9.00,39.00,35.75),
    array(5.00,8.00,37.00,32.75),
    array(3.00,4.00,24.00,24.75),

```

```
array(3.00,4.00,20.00,23.50),
array(3.00,10.00,43.00,34.50),
array(4.00,9.00,36.00,26.75),
array(5.00,10.00,42.50,36.50),
array(4.00,6.00,36.00,29.25),
array(5.00,8.00,39.00,31.25),
array(5.00,8.00,41.00,31.25)
```

Page | 278);

```
$centroid = array(
array(4.00,9.00,36.00,26.75),
array(4.00,9.00,36.00,19.25),
array(3.00,4.00,24.00,24.75),
array(3.00,4.00,20.00,23.50),
array(5.00,10.00,42.50,36.50)
);
```

IV. HASIL DAN PEMBAHASAN

Eksekusi program dengan output tabel per iterasi, berikut pengelompokan cluster nya di tiap-tiap iterasi.

TABEL X
HASIL AKHIR CLUSTERING

Ke-	K K	SL	RP S	RP P	C 1	C 2	C 3	C 4	C 5
Ke 1	4.00	9.00	36.00	26.75		X			
Ke 2	3.00	4.00	20.00	23.50				X	
Ke 3	4.00	9.00	36.00	19.25					X
Ke 4	5.00	6.00	35.00	39.00	X				
Ke 5	3.00	6.00	35.00	25.50		X			
Ke 6	5.00	9.00	39.00	35.75	X				
Ke 7	5.00	8.00	37.00	32.75	X				
Ke 8	3.00	4.00	24.00	24.75					X
Ke 9	3.00	4.00	20.00	23.50				X	
Ke 10	3.00	10.00	43.00	34.50	X				
Ke 11	4.00	9.00	36.00	26.75		X			
Ke 12	5.00	10.00	42.50	36.50	X				
Ke 13	4.00	6.00	36.00	29.25		X			
Ke 14	5.00	8.00	39.00	31.25	X				
Ke 15	5.00	8.00	41.00	31.25	X				

Hasil eksekusi program juga menampilkan nilai jarak yang dijadikan untuk penentuan kriteria cluster.

TABEL XI
JARAK CLUSTER PENENTUAN KATEORI

	KK	SL	RPS	RPP	Σ	Kat
Cluster 1	4.71	8.40	39.50	34.42	87.03	A
Cluster 2	3.75	7.50	35.75	27.06	74.06	B
Cluster 3	4.00	9.00	36.00	19.25	68.25	C

Cluster 4	3.00	4.00	24.00	24.75	55.75	D
Cluster 5	3.30	5.60	25.33	20.08	54.31	E

Dari tabel diatas dapat dijelaskan jika Σ dari cluster data mendekati dengan pola tabel diatas maka kan masuk kedalam kelompok cluster tersebut.

Tahapan selanjutnya adalah melakukan perbandingan antara data real dengan data hasil klasterisasi menggunakan k-means clustering, yang bertujuan untuk melihat tingkat akurasi dari algoritma terhadap 15 data dan 5 centroid cluster dari data penilai RPS dosen.

Proses perbandingan dimulai dengan membandingkan nilai kategori dari data real dengan hasil cluster dengan k-means dapat dilihat dari tabel berikut.

TABEL XII
PERBANDINGAN ANTAR DATA

Dosen	K-Means	Real	Hasil Ketepatan
Ke 1	B	B	Y
Ke 2	D	E	N
Ke 3	E	C	N
Ke 4	A	A	Y
Ke 5	B	C	N
Ke 6	A	A	Y
Ke 7	A	B	N
Ke 8	E	D	N
Ke 9	D	E	N
Ke 10	A	A	Y
Ke 11	B	B	Y
Ke 12	A	A	Y
Ke 13	B	B	Y
Ke 14	A	B	N
Ke 15	A	A	Y

Dari hasil perbandingan diatas ditemukan selisih perbandingan dengan jumlah Y = 8 dan N = 7 dari 15 total data dan jumlah 5 cluster centroid.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang didapat, maka penulis mengambil kesimpulan :

1. Metode *K-Means Clustering* adalah metode simpel yang dapat digunakan untuk mengklasterisasi data dalam jumlah banyak.

2. Bahwa jumlah ketepatan prediksi yang dilakukan oleh *algortima K-means* terhadap 15 data mengalami perbedaan sebanyak 53.33 % keakuratan prediksi.
3. Metode *k-means clustering* adalah salah satu teknik dari data mining, sehingga penulis menyimpulkan jika semakin banyak data yang di proses maka hasil juga akan semakin baik.

Page | 279

Berdasarkan kesimpulan diatas penulis dapat memberikan saran sebagai berikut :

1. Metode *k-means clustering* dengan pemrograman web masih membutuhkan optimasi yang lebih dari segi logic dan masalah *array* dalam bentuk nilai decimal panjang.
2. Menggukun data yang lebih banyak untuk melakukan test terhadap *k-means* untuk hasil prediksi yang lebih baik.
3. Algoritma *k-means* melakukan pembentukan *centorid* bisa di lakukan dengan memilih nilai centroid yang mendekati kategori Clusternya sehingga hasil *summary auclidian distance* bisa lebih baik.

REFERENSI

- [1] K. Monika, N. Lal and S. Qamar, "K-mean clustering algorithm approach for data mining of heterogeneous data," in *Information and Communication Technology for Sustainable Development*, vol. 10, Singapore, Springer, 2018, pp. 61-70.
- [2] R. Muliono, "Implementasi Algoritma Apriori Pada Data Benchmark Kosarak Dan Mushrooms," *JITE : JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 1, no. 1, pp. 34-41, September 2017.
- [3] T. Rismawan and K. Sri, "K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) dan Ukuran Kerangka," in *Seminar Nasional Aplikasi Teknologi Informasi 2008 (SNATI 2008)*, Yogyakarta, 2008.
- [4] H. Priyatman, F. Sajid and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 1, p. 2019.
- [5] F. N. R. F. J. Aziz, B. D. Setiawan and I. Arwani, "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 6, pp. 2243-2251, Juni 2018.
- [6] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 6, no. 4, pp. 161-168, Oktober 2018.
- [7] E. Rimmy and Chuchra, "Use of Data Mining Techniques for the Evaluation of Student Performance:A Case Study," *International Journal of Computer Science and Management Research*, vol. 1, no. 3, October 2012.
- [8] S. Budi, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.
- [9] M. E and K. S, "Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang," *Jurnal Bianglala Informatika*, vol. 3, no. 1, 2015.
- [10] P. Sari, B. Pramono and L. O. H. S. Sagala, "IMPROVE K-MEANS TERHADAP STATUS NILAI GIZI PADA BALITA," *SemanTIK*, vol. 3, no. 1, pp. 143-148, 2017.
- [11] LP2MP-UMA, *Panduan Kurikulum Pendidikan Tinggi Universitas Medan Area*, 2016.