

PENERAPAN METODE CLUSTERING K-MEANS UNTUK PROSES SELEKSI CALON PESERTA LOMBA MTQ

Ade Tria Rahmayanti Saragih, Abdul Sani Sembiring, Muhammad Sayuthi

Teknik Informatika STMIK Budi Darma, Medan, Indonesia
Jalan Sisingamangaraja No. 338 Simpang Limun, Medan, Indonesia

ABSTRAK

Dalam menyambut bulan suci Ramadhan walikota Tebing Tinggi mengadakan perlombaan-perlombaan. Salah satunya perlombaan membaca Al-Qur'an. Perlombaan ini bertujuan sebagai sarana menjaga kemurnian ayat-ayat Al-Qur'an melalui penghafal Al-Qur'an, serta menjaga ekstensi dari masa ke masa dengan menjaga tradisi lisan (pengucapan). Dalam penyeleksian peserta lomba sering sekali terjadi kecurangan dari pihak tertentu. Masalah yang sering terjadi memanipulasikan hasil nilai peserta. Data mining tidak lepas dari perkembangan teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi sejalan dengan pertumbuhan teknologi informasi. Dengan meningkatnya kapabilitas media penyimpanan yang semakin besar memungkinkan terjadinya akumulasi data dalam jumlah besar. Akumulasi data ini membentuk sebuah gudang data yang penting. Jika tidak digali, maka informasi penting yang ada akan menjadi tidak berguna. Tujuan penelitian ini adalah untuk menerapkan Clustering K-Means menjadi salah satu metode pengelompokan data nonhierarki (sekatan) untuk mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam suatu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain.

Kata Kunci: Pondok Pesantren Al-Hasyimiyah, MTQ, Clustering K-Means

I. PENDAHULUAN

Dalam menyambut bulan suci Ramadhan walikota Tebing Tinggi mengadakan perlombaan-perlombaan. Salah satunya perlombaan membaca Al-Qur'an. Perlombaan ini bertujuan sebagai sarana menjaga kemurnian ayat-ayat Al-Qur'an melalui penghafal Al-Qur'an, serta menjaga ekstensi dari masa ke masa dengan menjaga tradisi lisan (pengucapan). Dalam penyeleksian peserta lomba sering sekali terjadi kecurangan dari pihak tertentu. Masalah yang sering terjadi memanipulasikan hasil nilai peserta. Maka dari data yang didapatkan, Santri akan di seleksi dan membuat menjadi beberapa grup. Penyeleksian ini dilakukan berdasarkan kriteria-kriteria yang telah ditentukan.

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan mesin pembelajaran untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [1].

Clustering K-Means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) untuk mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam suatu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain [2]. Dalam penelitian sebelumnya pada jurnal *Simetris* yang berjudul "Klasterisasi proses seleksi pemain menggunakan algoritma K-means" (Narwati, 2014) menunjukkan bahwa penerapan *datamining* ada pada bidang penyeleksian, pengelompokan tersebut juga

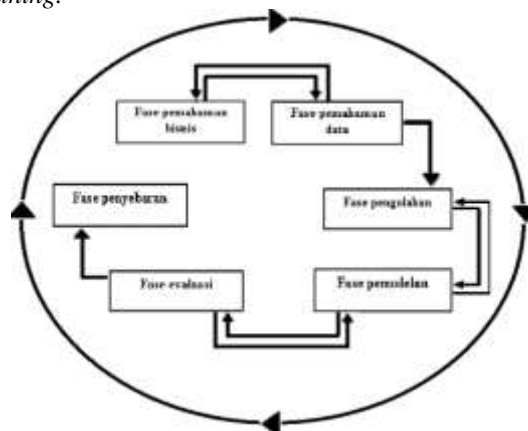
bisa di terapkan pada pembagian seleksi calon peserta perlombaan.

II. TEORITIS

A. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan mesin pembelajaran untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2].

Siklus hidup proyek *data mining* menurut *Cross-Industry Standart Proses for Data Mining (CRISP-DM)* yang dikembangkan tahun 1996 terbagi dalam 6 fase. Berikut gambar dari Siklus hidup proyek *data mining*.



Gambar 1. Proses siklus hidup data mining menurut CRISP-DM

Sumber: Kusrini, Emha Taufiq Luthfi, "Algoritma Data Mining" 2010 : 10 [2]

B. Clustering

Secara umum clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*[1].

C. Algoritma K-Means Clustering

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil [1]. langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut:

- Pilih jumlah *cluster k*.
- Inisialisasi *k* pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberiduberi nilai awal dengan angka-angka random.
- Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(x_{1i} - y_{1j})^2 + (x_{2i} - y_{2j})^2 + \dots + (x_{ki} - y_{kj})^2}$$

Dimana
 $D(i, j)$ = Jarak data ke i ke pusat cluster j
 X_{ki} = Data ke i pada atribut data ke k
 X_{kj} = Titik pusat ke j pada atribut ke k
- Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata – rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata – rata (mean) bukan satu – satunya ukuran yang bisa di pakai.
- Tugaskan lagi setiap objek memakai *cluster* yang bary. Jika pusat *cluster* tidak berubah lagi maka

proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

D. Rapid Miner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis pengelompokan. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksidalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining, termasuk operator untuk *input*, *output*, *datapreprocessing* dan visualisasi. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. *RapidMiner* ditulis dengan menggunakan bahasa *java* sehingga dapat bekerja dsemua sistem operasi[5].

III. ANALISA DAN PEMBAHASAN

Dalam menyambut bulan suci Ramadhan ini banyak dilakukan kegiatan perlombaan seperti nasyid, Musabaqoh Tilatiwati Qur'an (MTQ), adzan, dakwan. Penilaian diawal perlombaan biasanya hanya menggunakan sistem manual dan tidak transparan. Maka dari itu, dalam penyeleksian peserta lomba sering sekali terjadi kecurangan dari pihak tertentu. Masalah yang sering terjadi memanipulasikan hasil nilai data seleksi peserta. Tidak hanya memanipulasikan data, kecurangan juga sering terjadi dari data santri yang mengikuti perlombaan. bebarapa Santri akan di seleksi untuk mewakili perlombaan MTQ tingkat kecamatan. Penyeleksian ini dilakukan berdasarkan kriteria-kriteria yang telah ditentukan..Berdasarkan hasil pengamatan penilaian di Pondok Pesantren Al-Hasyimiyah ini diketahui belum maksimal dan efisien maka perlu dilakukan penelitian penyeleksian para santri yang akan mewakili perlombaan ketingkat Kecamatan.

A. Data Calon Peserta Lomba

Data-data yang menjadi pokok pembahasan dalam penelitian ini adalah data-data santri yang ingin mengikuti lomba MTQ pada tahun 2015, sebanyak 84 data, yang meliputi santri dari umur 13thn-15thn berjumlah 23 santri dengan golongan maksimal menguasai 5 Juz dan santri umur 16-18thn berjumlah 61 santri dengan golongan 10 Juz. Berdasarkan tabel berikut:

Tabel 1. Kriteria Dari Juz

Golongan	Nilai	Kriteria
≤ 3 Juz	0 - 59	Kurang (Gagal)
≤ 6 Juz	60 -80	Baik
≤ 10 Juz	81 - 100	Sangat Baik

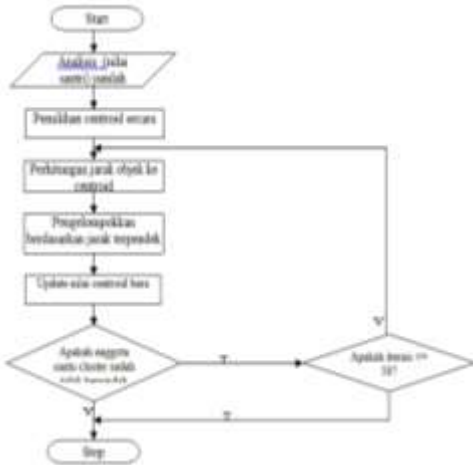
Sesuai dengan kriteria dari hafalan Juz yang telah ditentukan, maka dari 84 peserta yang lolos seleksi

hanya 73 peserta yang akan di transformasikan sesuai dengan algoritma *K-means clustering*.

B. Penerapan Algoritma K-Means Clustering

Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil[4]. Algoritma K-Means akan melakukan pengulangan langkah sampai tidak ada obyek yang dapat berpindah :

1. Menentukan koordinat titik tengah setiap cluster,
2. Menentukan jarak setiap obyek terhadap koordinat titik tengah,
3. Mengelompokkan obyek tersebut berdasarkan pada jarak minimumnya.



Gambar 2. Flowchart Algoritma K-Means Clustering

Berdasarkan data hasil penyeleksian di dapatkan hasil nilai santri . Berikut ini nilai hasil seleksi santri sebagai berikut:

Tabel 2. Pola Data Nilai Peserta Pada Pondok Pesantren Al-Hasyimiyah

No	Nama Santri	Jenis Kelamin	Usia	Nilai Tajwid	Nilai Penulisan	Nilai Suara
1	M. Fauzan Srgh	L	15	88	62	C
2	Ahmad Sayuti	L	16	87	75	B
3	Ayu Sri Lestari	P	15	78	80	A
4	Irfan Syahputra	L	16	90	70	C
5	Aditya W.	L	18	76	70	C
6	Sari Mutiara	P	18	78	75	B
7	Alya Kirana S.	P	17	94	85	B
8	Ade Rizal Srgh	L	16	82	90	B
9	Bunga Pratiwi	P	13	89	63	B
10	Indah Maya Sari	P	17	93	65	B
11	Rizky Syahputra	L	17	92	77	A
12	M. Reza Sinaga	L	17	84	76	C
13	Siti Aisyah	P	18	90	68	B

14	Siti Rohma	P	16	88	69	C
15	Yudha Rinaldi T.	L	18	88	85	A
16	Abdi Subhan	L	17	87	76	C
17	Ulfa Kurnia Rahma	P	15	90	87	B
18	Zulfan Effendi Srgh	L	14	77	65	C
19	Indah Sari	P	14	81	61	C
20	Zaini Abdullah	L	14	90	70	A
21	Ditta P.Sari	P	18	87	80	B
22	Khadijah	P	18	95	85	A
23	Cut Mariska	P	18	87	77	B
24	Widya Anggreini	P	13	93	87	D
25	Wulan Suci Tri Sakti	P	14	87	90	B
26	Dwi Sartika	P	18	79	86	B
27	Nanda Rusmana	L	18	89	80	B
28	Rafika Putri Mandasari	P	18	81	66	C
29	Leni Hadiananti	P	17	87	75	B
30	Herry Hardiansyah	L	17	75	75	C
31	Ridho Ardiansyah	L	16	88	65	C
32	M. Isnin Alamsyah	L	13	87	68	D
33	Dewi Syahfitri	P	15	79	85	B
34	Nanda Ira Ayu	P	16	91	75	A
35	Umi Kalsum	P	14	81	65	C
36	Nuralila	P	15	87	78	B
37	Farhat Ramdhan	L	14	77	76	B
38	Alwi Lesmana	L	17	96	83	A
39	Utari nidya	P	18	75	68	C
40	Nur Hidayah	P	17	79	78	B
41	Teguh santoko	L	17	88	68	B
42	Imran	L	16	80	69	B
43	Mentari Nuriatsy	P	15	80	74	B
44	Khairin Nasroh	P	16	87	65	C
45	Tika wulansari	P	17	93	66	B
46	Bobby Lesmana	L	18	79	76	A
47	Dedek Dermawan	L	17	78	70	C
48	Ade Sri Lili Ani	P	17	85	65	B
49	Sultan Hasim	L	14	88	65	B
50	Aditya Rizky	L	15	80	66	A
51	Wulan Nur F.	P	16	84	78	B
52	Teuku Wisnu	L	17	75	67	C
53	Aulia Husni	L	15	95	85	A
54	Hamzah R.	L	18	83	76	B
55	Nabila Tifani Nst	P	17	79	77	B
56	Hendra Winata	L	17	93	69	C
57	Satria Fauzan	L	16	81	67	C
58	Ulzannah Zein Nst	P	16	83	77	B
59	Rizki Ananda T	P	17	93	78	B
60	Rohana P.	P	17	96	80	A
61	Sarmutiara	P	17	94	82	A
62	Santi	P	16	88	84	A
63	Sartini Hamid	P	15	81	75	B
64	Sri Dewi M.	P	15	93	67	C
65	Sundari	P	14	91	84	B
66	Ulfah Kurnia R.	P	14	86	85	B
67	Vika Andriani	P	17	90	77	B
68	Winda Sari	P	18	83	78	C

69	Yuda Rinaldi T	L	15	87	80	B
70	Yusril Hija Mahendra	L	15	90	82	A
71	Zaini Abdullah	L	14	80	84	B
72	Zono Basuki	L	14	82	75	B
73	Zulpan Efendi Srg	L	17	75	67	C

Untuk melakukan inisialisasi santri yang lulus tahap awal dilakukan dengan langkah-langkah sebagai berikut :

1. Pada data nilai hasil seleksi terlebih dahulu dilakukan pembagian hasil nilai yang menjadi beberapa bagian nilai.
2. Kemudian Nilai-nilai tersebut diurutkan dari yang terbesar berdasarkan frekuensi santri yang di dapat pada nilai hasil seleksi.
3. Setelah itu nilai yang memiliki frekuensi terbesar diberi inisial 1 dan nilai yang memiliki frekuensi terbesar kedua diberi inisial dengan angka, begitu seterusnya hingga nilai dengan frekuensi paling sedikit.

Untuk dapat melakukan pengelompokkan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu :

1. Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data yang akan dikelompokkan menjadi dua *cluster*.
2. Tentukan titik pusat awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel 3.3:

Cluster 1 : di ambil dari data ke-27

Cluster 2 : di ambil dari data ke-54

Cluster 3 : di ambil dari data ke-24

3. Tempatkan setiap data pada *cluster*. Dalam penelitian ini digunakan metode *hard k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data santri pertama ke pusat *cluster* pertama, menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$\begin{aligned}
 d(1,1) &= \sqrt{(1-1)^2 + (1-1)^2 + (2-1)^2} \\
 &= \sqrt{(0) + (0) + (1)} \\
 &= \sqrt{1} \\
 d(1,1) &= 1
 \end{aligned}$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data santri pertama dengan *cluster* pertama adalah 1

Jarak dari data santri pertama ke pusat *cluster* kedua:

$$\begin{aligned}
 d(1,2) &= \sqrt{(1-3)^2 + (1-1)^2 + (2-1)^2} \\
 &= \sqrt{(4) + (0) + (1)} \\
 &= \sqrt{5} \\
 d(1,2) &= 2,236
 \end{aligned}$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data santri pertama dengan pusat *cluster* kedua adalah 2,236.

Dalam penelitian ini, iterasi *clustering* data santri terjadi sebanyak 3 kali iterasi. Pada iterasi ke-3 ini, titik pusat dari setiap *cluster* yang tidak berubah dan tidak ada data yang berpindah antar *cluster*.

Dari hasil *cluster* 1 memiliki pusat (1,413, 1,195, 1,695) adapun yang masuk pada *cluster* 1 yaitu data ke- (1, 2, 3, 4, 5, 9, 13, 14, 15, 16, 17, 18, 21, 23, 25, 27, 29, 30, 31, 32, 36, 39, 41, 44, 46, 47, 49, 50, 52, 62, 65, 67, 69, 70, 73). Terlihat bahwa karakteristik siswa pada *cluster* 1 didominasi oleh nilai suara berkisar 87-92, berdasarkan nilai seleksi santri didominasi oleh santri yang memperoleh nilai berkisar dari 60-80, berdasarkan nilai huruf di dominasi oleh santri yang memiliki nilai B.

Dari hasil *cluster* 2 memiliki pusat (2,658, 1,219, 1,585) adapun yang masuk pada *cluster* 2 yaitu data ke- (6, 7, 8, 10, 12, 19, 26, 28, 33, 35, 37, 40, 42, 43, 45, 48, 51, 54, 55, 56, 57, 58, 59, 63, 66, 68, 71, 72). Terlihat bahwa karakteristik pada *cluster* 2 didominasi oleh nilai suara berkisar 81-86, berdasarkan nilai seleksi santri didominasi dengan nilai berkisar dari 60- 80, berdasarkan nilai huruf santri di dominasi oleh nilai B.

Dari hasil *cluster* 3 memiliki pusat (4, 1.8, 3) adapun yang masuk pada *cluster* 3 yaitu data ke- (11, 20, 22, 24, 34, 38, 53, 60, 61, 64). Terlihat bahwa

Titik Pusat Awal	Nama Santri	Nilai Suara	Nilai Seleksi	Nilai
Cluster 1	Nanda Rusmana	1	1	1
Cluster 2	Hamzah Ramadhan	3	1	1
Cluster 3	Widya Anggreini	4	2	4

karakteristik santri pada *cluster* 3 didominasi oleh nilai suara berkisar 93-100, berdasarkan nilai seleksi santri di dominasi nilai 80-100, berdasarkan nilai huruf santri di dominasi nilai A.

Tabel 3. Hasil seleksi calon peserta lomba

No	Nama	Jenis Kelamin		Umur	Golongan
		P	L		
1	Bunga Pratiwi	✓		13 Thn	Remaja
2	Ulfa Kurnia Rahma	✓		15 Thn	Remaja
3	Indah Sari	✓		14 Thn	Remaja
4	Zulfikar Insani		✓	17 Thn	Dewasa

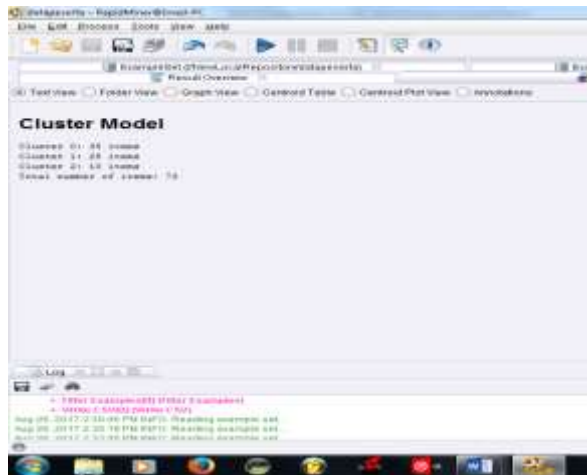
5	Khalifah Nur Isma	✓		17 thn	Dewasa
6	M. Isnin Alamsyah		✓	15 Thn	Remaja
7	Khalifatul rizki	✓		17 Thn	Dewasa
8	Dimas agung Widodo		✓	17 thn	Dewasa
9	Aditya Rizky		✓	15 Thn	Remaja
10	Aulia Husni		✓	15 thn	Remaja

IV. IMPLEMENTASI

Tahap implementasi sistem merupakan tahap untuk mengaplikasikan apa hasil analisa. Implementasi ini dilakukan sebagai wujud nyata hasil analisa ke dalam bahasa pemrograman untuk diterapkan pada keadaan yang sesungguhnya. Hasil dari penerapan tersebut diharapkan dapat siap di uji dan digunakan.

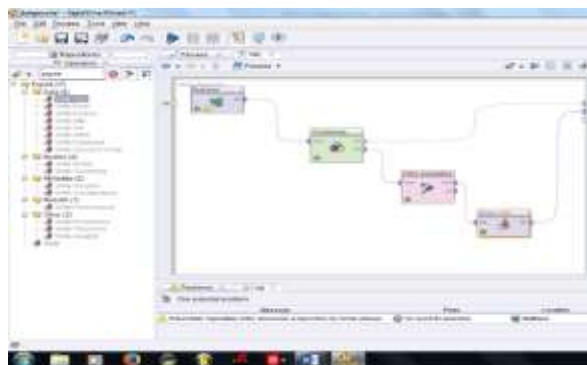
Tampilan Result Overview

Pengujian menggunakan aplikasi *rapidminer*, dan berikut adalah tampilannya:



Gambar 3. Tampilan Result Overview

Dimana ada 3 *clustermodel*, dengan jumlah total 73 *item*. Berikut ini adalah pengolahan data dengan menggunakan *k-means clustering* pada *rapidminer*.



Gambar 4. Pemodelan K-Means Clustering pada Rapidminer

Dengan menggunakan pemodelan *k-means clustering* seperti gambar 4.2 di atas, dengan jumlah *cluster* sebanyak 3 buah, maka didapatkan hasil dengan *cluster* yang terbentuk adalah 3, sesuai dengan pendefinisian nilai *k* dengan jumlah *cluster_0* ada 35 *items*, *cluster_1* ada 28 *items*, *cluster_2* ada 10 *items*, dengan total sejumlah 36 *items*.

Tabel 4. Hasil cluster k-means cluster dalam implementasi rapidminer

Cluster Model	
Cluster_0	35 items
Cluster_1	28items
Cluster_2	10 items
Total number of items	73 items

V. KESIMPULAN

Berdasarkan uraian implementasi algoritma *K-Means Clustering*, maka penulis dapat memberikan kesimpulan bahwa:

1. Penyeleksian calon peserta yang mewakili belum efisien dan efektif. Masih banyak kecurangan memanipulasi nilai dari pihak tertentu. Penilaian kurang transparan.
2. Melakukan penyeleksian calon peserta berdasarkan data-data nilai tajwid, nilai penjiwaan, nilai seleksi penguasaan juz dan nilai suara yang paling banyak didominasi santri kemudian akan dikategorikan menjadi santri yang akan mewakili untuk perlombaan.
3. Algoritma *K-Means Clustering* pada *rapidminer* dimulai dengan penginputan data nilai santri menjadi database pada *Ms.Excel* dengan ekstensi *csv*. Mengelompokkan data dengan algoritma *k-measn clustering*, dilakukan dengan cara menentukan jumlah *cluster*, hitung jarak terdekat dengan pusat *cluster*. Data dengan jarak terdekat menyatakan anggota *cluster* tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada *cluster* lain, untuk meminimalkan fungsi objektif.

REFERENCES

- [1] Eko Prasetyo, "Pengertian Data Mining dan bukan Data Mining", in Data mining, Konsep dan Aplikasi menggunakan Matlab", Yogyakarta, Andi Soffet, (2010), pp 2-3
- [2] Kusriani, Emha Taufiq Luthfi, "Bab II Data Mining", in Algoritma Data Mining, Yogyakarta, Andi Offset, 2010, pp 3-12.
- [3] Denis Aprilia C, Donny Aji Baskoro, Lia Ambarwati, I Wayan Simri Wicaksana, "Belajar Data Mining Dengan RapidMiner", 2013.
- [4] Pondok Pesantren Al-Hasyimiyah, "Logo Ponpes", 2013, pp 14
- [5] Denis Aprilia C, Donny Aji Baskoro, Lia Ambarwati, I Wayan Simri Wicaksana, "Belajar Data Mining Dengan RapidMiner", 2013.

- [6] E. Buulolo, N. Silalahi, Fadlina, and R. Rahim, "C4.5 Algorithm To Predict the Impact of the Earthquake," *Int. J. Eng. Res. Technol.*, vol. 6, no. 2, 2017.
- [7] H. Widayu, S. D. Nasution, N. Silalahi, and M. Mesran, "DATA MINING UNTUK MEMPREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN ALGORITMA C4.5," *MEDIA Inform. BUDIDARMA*, vol. 1, no. 2, Jun. 2017.
- [8] E. Buulolo, "ALGORITMA APRIORI PADA DATA PENJUALAN DI SUPERMARKET," in *Seminar Nasional Inovasi dan Teknologi Informasi 2015 (SNITI)*, 2015, no. September 2015, pp. 4–7.
- [9] J. Simarmata, *Pengenalan Teknologi Komputer dan Informasi*. Yogyakarta: Andi, 2006.
- [10] M. Maharani et al., "IMPLEMENTASI DATA MINING UNTUK PENGATURAN LAYOUT MINIMARKET DENGAN MENERAPKAN ASSOCIATION RULE," *J. Ris. Komput.*, vol. 4, no. 4, pp. 6–11, 2017.
- [11] J. Simarmata, *Rekayasa Perangkat Lunak*. Yogyakarta: Andi, 2010.
- [12] F. T. Waruwu, E. Buulolo, and E. Ndruru, "IMPLEMENTASI ALGORITMA APRIORI PADA ANALISA POLA DATA PENYAKIT MANUSIA YANG DISEBABKAN OLEH ROKOK," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. I, no. 1, pp. 176–182, 2017.
- [13] E. Buulolo, "Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus : Apotik Rumah Sakit Estomihi Medan)," *Pelita Inform. Budi Dharma*, vol. 4, no. Agustus 2013, pp. 71–83, 2013.
- [14] W. Fitriani and A. P. U. Siahaan, "Comparison Between WEKA and Salford System in Data Mining Software," *Int. J. Mob. Comput. Appl.*, vol. 3, no. 4, pp. 1–4, 2016.
- [15] Y. A. Fiandra, S. Defit, and Yuhandri, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 2, pp. 82–89, 2017.
- [16] L. Marlina, M. Muslim, and A. P. U. Siahaan, "Data Mining Classification Comparison (Naïve Bayes and C4 . 5 Algorithms)," *Int. J. Eng. Trends Technol.*, vol. 38, no. 7, pp. 380–383, 2016.