

Data Mining Clustering Nilai IQ Siswa Sekolah SMAK Yos Sudarso Batam

Alvendo Wahyu Aranski *, Koko Handoko

Universitas Putera Batam, Batam

* kokohandoko01@gmail.com

Abstract

Yos Sudarso School Batam is a private school most widely known by the Batam community, the school managed to get first place, in the average score of the Riau Islands national exam. Grouping students based on IQ levels. So that students can accept more appropriate and effective learning methods that can be accepted by students with lessons. At the Yos Sudarso High School in Batam which is a place of research, students who take part in school very much and grouping students will collect a lot of time and are not efficient if done manually. Grouping students can be done quickly and precisely using data mining. One method in data mining is grouping or clustering. The stages of this method begin with random selection K, K here is a collection to be formed. Then assign K values randomly, for a while this value becomes the center of the cluster or commonly called centroid, meaning or "meaning". Calculate the distance of each data in each centroid using the Euclidian formula until the closest distance from each data is found with a centroid after it has been manually calculated using the data mining software RapidMiner. The method of using K-means to do grouping is one of the proper methods that can be seen from the variables to be used is the value of students' IQ levels.

Keywords: *Data Mining; K-Means; Level of IQ.*

Abstrak

Sekolah Yos Sudarso Batam merupakan suatu sekolah swasta yang banyak diketahui oleh masyarakat batam, sekolah berhasil memperoleh peringkat pertama, dalam rata-rata nilai ujian nasional tingkat Kepulauan Riau. Pengelompokan siswa berdasarkan tingkat IQ bertujuan agar siswa dapat menerima metode pembelajaran yang lebih tepat dan efektif sehingga siswa dapat menerima pelajaran dengan baik. Pada sekolah SMAK Yos Sudarso batam yang menjadi tempat penelitian, siswa yang mendaftarkan diri untuk bersekolah sangat banyak dan pengelompokan siswa akan memakan waktu yang sangat banyak dan tidak efisien jika dilakukan secara manual. Pengelompokan siswa dapat dilakukan secara cepat dan tepat dengan memanfaatkan *data mining*. Salah satu metode pada *data mining* adalah pengelompokan atau *clustering*. Tahapan metode ini dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, mean atau "means". Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid* setelah selesai perhitungan secara manual di jalankan dengan *software* data mining yaitu *RapidMiner*. Penggunaan metode *K-Means* untuk melakukan pengelompokan merupakan salah satu metode yang tepat jika dilihat dari variabel yang akan digunakan yaitu nilai tingkat IQ siswa.

Kata Kunci: *Data Mining; Metode K-means Clustering; Tingkat IQ.*

1. Pendahuluan

Pada sekolah SMAK Yos Sudarso Batam, setiap siswa yang masuk akan dilakukan tes IQ. Disebabkan banyaknya siswa yang mendaftar, terkadang nilai-nilai IQ siswa menjadi data yang terlalu banyak dan menyebabkan kesulitan untuk guru dalam memprosesnya. Klasifikasi yang dilakukan secara manual terhadap data sebanyak itu dapat menyebabkan ketidaktepatan pada prosesnya. Proses ini dapat dibantu dengan menggunakan *data*

mining. Beberapa peneliti menghubungkan bahwa manusia dengan tingkatan IQ yang lebih tinggi cenderung lebih cerdas dibandingkan manusia dengan tingkatan IQ yang lebih rendah. Pengelompokan siswa berdasarkan tingkatan IQ sangat diperlukan dalam melakukan pengajaran yang tepat terhadap siswa. Siswa dengan tingkatan IQ yang lebih rendah membutuhkan perhatian lebih agar dapat mengikuti pelajaran dengan baik. *Data mining* adalah proses yang menggunakan

teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar (Masalah, 2011). Banyak metode yang dimiliki *data mining* sesuai dengan kebutuhan dan masalah yang dihadapi. Pada kasus ini, metode yang dibutuhkan adalah pengelompokan. Salah satu algoritma yang sering digunakan dalam melakukan pengelompokan adalah algoritma *K-means*. *K-means* merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok.

2. Kajian Literatur

Beberapa teori yang peneliti gunakan dalam pembuatan dan penyelesaian masalah ini adalah:

2.1 Knowledge Discovery in Database

Konsep *data mining* memiliki beberapa teknik pendekatan yaitu secara kuantitatif dan kualitatif. Pendekatan kuantitatif terdiri dari pendekatan probabilitas seperti logika induktif, pencarian kesamaan pola serta pohon keputusan. Pendekatan kualitatif meliputi pendekatan deviasi, analisa kecenderungan, algoritma genetik serta jaringan saraf tiruan. Pada umumnya, terdapat beberapa ciri penting dalam sebuah pencarian pengetahuan dalam proses KDD, yaitu:

- (1) Data yang dikerjakan dalam jumlah yang besar.
- (2) Volume data harus diefisienkan.
- (3) Memprioritaskan akurasi.
- (4) Bahasa yang digunakan merupakan bahasa tingkat tinggi.
- (5) Menggunakan beberapa bentuk dari pembelajaran otomatis.
- (6) Menghasilkan hasil berupa pengetahuan yang dapat bermanfaat.

2.2 Data Mining

Data mining juga dapat didefinisikan sebagai suatu rangkaian proses dengan tujuan untuk menggali sebuah nilai tambah dari sebuah kumpulan data yang berupa pengetahuan yang selama ini tidak dapat diperoleh dengan cara penggalian manual. *Data mining* sendiri merupakan salah cabang dari bidang ilmu *artificial intelligence* dan *machine learning* yang digabungkan dengan statistik dan matematika yang memanfaatkan basis data (Handoko, 2016).

Data mining merupakan serangkaian proses yang dibagi menjadi beberapa tahap. Tahap-tahap tersebut saling berhubungan yang dimana *user* terlibat langsung atau dengan perantara pada *knowledge base*. Tahap-tahap ini diilustrasikan di Gambar (Bisilisin,

Herdiyeni, & Silalahi, n.d.):

- (1) Pembersihan data (untuk membuang data yang tidak konsisten dan noise)
- (2) Integrasi data (penggabungan data dari beberapa sumber)
- (3) Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining)
- (4) Aplikasi teknik DM
- (5) Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai)
- (6) Presentasi pengetahuan (dengan teknik visualisasi)

2.3 K-Means

K-Means Clustering merupakan metode yang termasuk ke dalam golongan algoritma *Partitioning Clustering*.

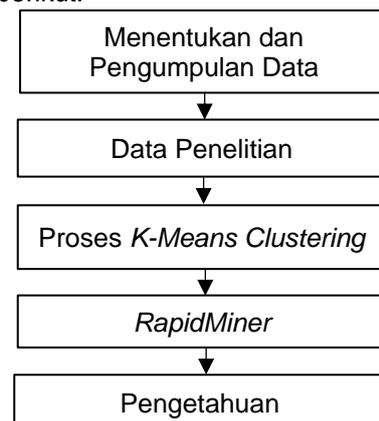
Langkah-langkah dari metode *K-Means* adalah sebagai berikut (Handoko, 2016):

- (1) Tentukan nilai *k* sebagai jumlah *cluster* yang ingin dibentuk.
- (2) Bangkitkan *k centroid* (titik pusat cluster) awal secara acak.
- (3) Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua objek (Euclidean Distance).
- (4) Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya.
- (5) Tentukan posisi *centroid* baru (kC) dengan cara menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama.

Sebelum algoritma *k-means* menyatu, perhitungan jarak dan sentroid dilakukan saat *loop* dieksekusi beberapa kali, katakanlah *l*, di mana bilangan bulat positif *l* dikenal sebagai jumlah iterasi *k-means*. Nilai yang tepat dari *l* bervariasi tergantung pada *centroid* awal awal bahkan pada dataset yang sama ("Application of *k-Means Clustering* algorithm for prediction of Students' Academic Performance," 2010).

3. Metode Penelitian

Metode penelitian yang digunakan adalah sebagai berikut:



Gambar 1. Metode Penelitian

4. Hasil dan Pembahasan

4.1 Hasil Penelitian

Untuk mendapatkan hasil penelitian yang baik tentang *data mining* maka diperlukan suatu pengelompokan nilai IQ siswa dengan metode *k-means clustering*. Metode ini dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, mean atau "means". Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*.

Proses *clustering* atau pengelompokan di atas maka dihasilkan pengelompokan ke dalam *cluster* sebagai berikut:

(1) Cluster 0

Setelah melakukan perhitungan secara manual sesuai dengan langkah-langkah yang di atas, maka dapat dikelompokkan ke dalam *cluster 0* seperti pada tabel berikut :

Tabel 1 Hasil Pengelompokan *Cluster 0*

NO	PR	PO	PB	PV
1	119	124	110	116
2	113	120	120	117
3	117	120	114	118
4	125	121	113	115
5	110	113	117	121
6	119	120	120	113
7	119	120	117	114
8	122	113	114	119
9	115	120	117	112
10	117	120	123	112
11	117	118	107	112
12	112	110	117	115
13	120	112	126	114
14	111	115	114	110
15	120	112	110	112
16	110	110	110	112
17	113	110	100	108
18	111	100	104	107

(2) Cluster 1

Setelah melakukan perhitungan secara manual sesuai dengan langkah-langkah yang di atas, maka dapat dikelompokkan ke dalam *cluster 1* seperti pada tabel berikut:

Tabel 2 Hasil Pengelompokan *Cluster 1*

NO	PR	PO	PB	PV
1	131	144	145	138
2	138	142	140	138
3	136	131	148	139
4	120	136	138	133
5	123	130	139	138
6	133	136	135	130
7	121	132	136	130
8	128	123	142	138
9	133	142	148	140
10	130	136	145	137
11	127	134	138	133
12	138	141	145	127
13	120	130	138	132

(3) Cluster 2

Setelah melakukan perhitungan secara manual sesuai dengan langkah-langkah yang di atas, maka dapat dikelompokkan ke dalam *cluster 2* seperti tabel berikut:

Tabel 3 Hasil Pengelompokan *Cluster 2*

NO	PR	PO	PB	PV
1	126	130	129	127
2	120	130	126	125
3	124	130	125	123
4	130	126	129	128
5	120	130	117	123
6	130	130	123	122
7	130	125	126	123
8	120	122	132	125
9	120	121	126	123
10	124	124	123	120
11	124	121	126	119
12	128	123	126	117
13	126	121	123	119
14	118	113	132	130
15	137	134	126	127
16	135	130	129	131
17	126	130	135	128
18	128	129	132	127
19	120	130	123	126

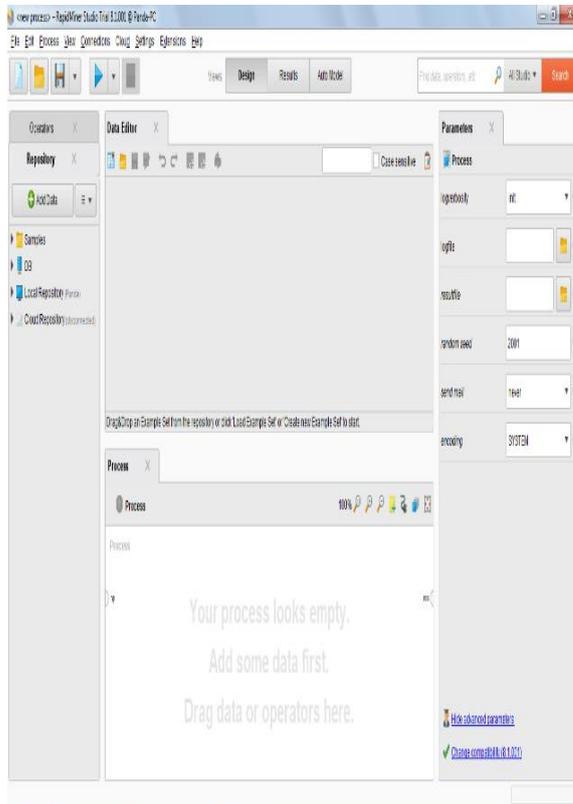
Dari kesimpulan pengelompokan data di atas dapat diketahui kelompok-kelompok nilai IQ siswa Sekolah SMAK Yos Sudarso Batam.

Untuk membuktikan kebenaran pada tahap analisa dan hasil secara manual maka perlu

dilakukan pengujian lagi untuk pengelompokan data IQ Siswa Sekolah akan menggunakan algoritma *K-Means* pada aplikasi pengujian *RapidMiner*. Adapun tahapan - tahapan yang dilakukan dalam pengujian ini yaitu :

- Tampilan *RapidMiner*

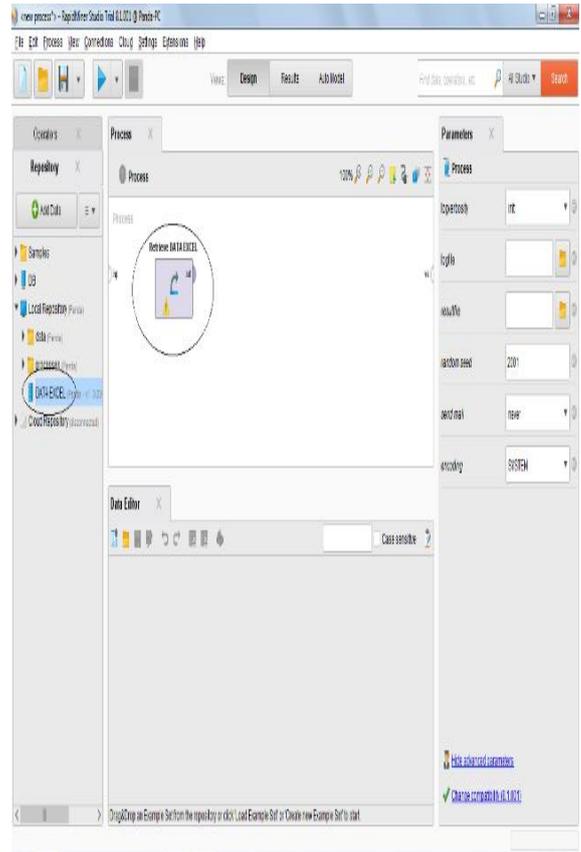
Pertama sekali pilih *New Process* lalu *Blank* pada halaman awal, lalu akan tampilan menu kosong pada aplikasi *RapidMiner* dapat dilihat pada gambar berikut :



Gambar 2. Tampilan *RapidMiner*

Dari tampilan *RapidMiner* lalu *import* data, pemilihan data *input*, dan sortir data.

- Proses *K-Means Clustering*



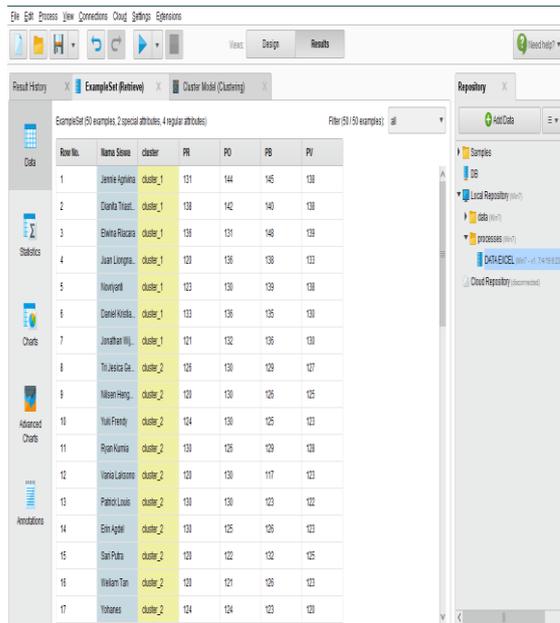
Gambar 3. Proses *K-Means*

Langkah selanjutnya proses pemindahan data, *input* metode *k-means*, proses perubahan nilai *k*, halaman akhir proses,

4.2 Pembahasan

Setelah IQ Siswa berhasil diproses, maka dapat dilihat hasil pengelompokan data. Hasil *clustering* juga dapat dilihat dalam format *charts* yang berupa titik-titik yang menyebar seperti gambar berikut:

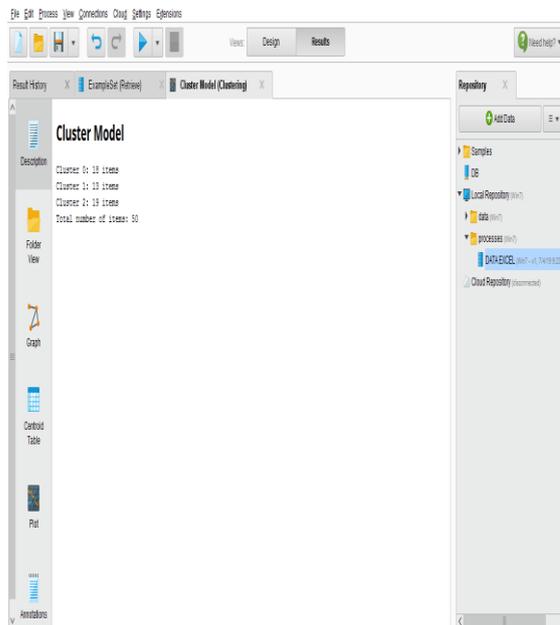
Selanjutnya setelah *data base* berhasil diproses, maka dapat dilihat hasil pengelompokan data. Pada *RapidMiner* versi 8.1, hasil proses dapat dilihat dalam banyak bentuk seperti pada gambar di bawah ini:



Row No.	Nama Siswa	cluster	PR	PO	PB	PV
1	Jennie Agnina	cluster_1	131	144	145	138
2	Diana Triast	cluster_1	138	142	140	138
3	Enrika Rizca	cluster_1	138	131	148	138
4	Juan Longra	cluster_1	133	138	138	133
5	Noviyanji	cluster_1	123	130	138	138
6	Daniel Krista	cluster_1	133	138	135	138
7	Jonathan Will	cluster_1	121	132	136	138
8	Tri Jeksa De	cluster_2	128	130	129	127
9	Milani Heng	cluster_2	120	130	126	125
10	Yuki Fendy	cluster_2	124	130	125	123
11	Ryan Nurma	cluster_2	130	128	129	128
12	Yanis Larsson	cluster_2	120	138	117	123
13	Patrick Louis	cluster_2	130	130	123	122
14	Elin Agnes	cluster_2	130	125	126	123
15	Sari Putra	cluster_2	120	122	132	125
16	Wahidun Tun	cluster_2	120	121	128	123
17	Yohanes	cluster_2	124	124	123	120

Gambar 4. Hasil Clustering Berupa Tabel

Selanjutnya hasil *clustering* dapat dilihat dengan tampilan *text view* seperti gambar berikut:

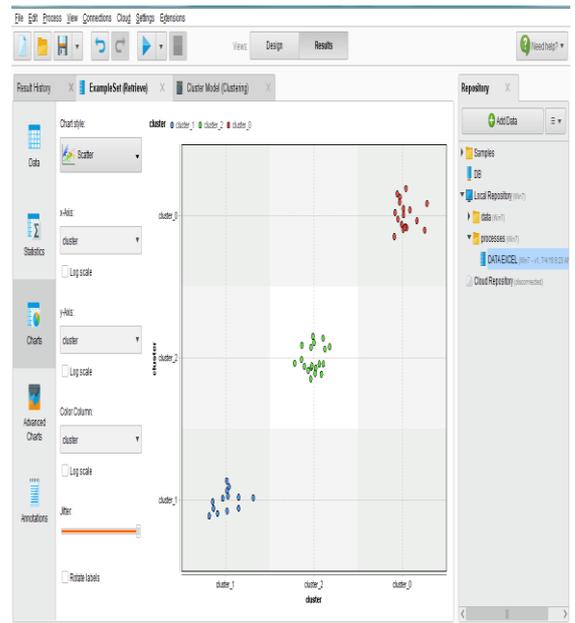


Cluster Model

Description
 Cluster 0: 18 items
 Cluster 1: 13 items
 Cluster 2: 19 items
 Total number of items: 50

Gambar 5. Hasil Pengelompokan

Hasil *clustering* juga dapat dilihat dalam format *charts* yang berupa titik-titik yang menyebar seperti gambar berikut:



Gambar 6. Hasil Pengelompokan

Dapat terlihat hasil pengelompokan nilai IQ Siswa Sekolah SMAK Yos Sudarso dari 50 sampel data.

5. Kesimpulan dan Saran

Adapun kesimpulan dari penelitian ini adalah:

Hasil Karakteristik yang dihasilkan dalam proses *clustering* adalah pada setiap *cluster* yang telah dihasilkan memiliki sifat yang hampir sama antara data-data yang berada dalam satu *cluster*, yaitu nilai IQ siswa, sehingga dengan melakukan karakteristik *clustering*, pengetahuan mengenai pengelompokan siswa berdasarkan nilai IQ dapat membantu sekolah SMAK Yos Sudarso Batam untuk menentukan pembelajaran yang tepat kepada siswa dengan tingkatan IQ yang sesuai.

Saran dari penelitian ini adalah penelitian selanjutnya bisa menerapkan algoritma-algoritma lainnya dan *software-software* yang digunakan untuk metode *clustering*.

Ucapan Terima Kasih

Ucapan terima kasih kepada DRPPM DIKTI yang sudah membiayai penelitian ini sehingga penelitian ini bisa terlaksana dengan baik.

Daftar Pustaka

Asringtias, Y., Mardiyah, R., Studi, P., Informatika, T., Bisnis, F., Informasi, T., & Yogyakarta, U. T. (2014). APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI, 8(1), 837-848.

Bisilisin, F. Y., Herdiyeni, Y., & Silalahi, B. I. B. P. (n.d.). Optimasi K-Means Clustering Menggunakan Particle Swarm Optimization pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra K-Means Clustering Optimization Using

- Particle Swarm Optimization on Image Based Medicinal Plant Identification System, 3(2002).
- Handoko, K. (2016). Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan). *Jurnal Teknologi Dan Sistem Informasi*, 2(3), 31
- Masalah, L. B. (2011). Data mining.
- Putra, R. R., & Wadisman, C. (2018). IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS, 1(1), 72–77.