



INFOMATEK

Volume 23 Nomor 2 Desember 2021

PENERAPAN ALGORITMA K-MEANS UNTUK MENENTUKAN KELAS UNGGULAN DI SMP PELITA BANDUNG

Novi*, Ade Mubarak

Program Studi Sistem Informasi
Universitas ARS

Abstrak: Kelas Unggulan pada SMP Pelita Bandung merupakan program yang dilakukan oleh pihak sekolah untuk meningkatkan kualitas siswanya. Kelas unggulan yang telah dibentuk dirasa masih kurang efektif sehingga kesulitan untuk menentukan nilai siswa-siswi untuk dimasukan ke kelas unggulan. Salah satu untuk mengatasi masalah yang ada adalah dengan menggunakan penerapan *Data Mining* yang bisa digunakan untuk pengolahan data menjadi sumber informasi strategis dan lebih mudah bagi sekolah menyeleksi siswa masuk kelas unggulan maka peneliti ini menggunakan metode Algoritma *K-Means Clustering*. Sebagai penerapan metode *clustering* untuk data perhitungan algoritma *K-Means* yang digunakan adalah nilai siswa dan masing-masing data nilai digunakan sebagai atribut. Data yang diambil yaitu 119 data yang dibagi menjadi 3 *cluster*, Pada *cluster* pertama dengan rata-rata terendah akan dimasukan ke dalam kelas C sebanyak 42 siswa, pada *cluster* kedua dengan rata-rata sedang akan dimasukan ke dalam kelas B sebanyak 37 siswa, sedangkan pada *cluster* ketiga dengan nilai rata-rata tertinggi akan dimasukan ke dalam kelas A sebanyak 40 siswa. Hasil dari penelitian ini menunjukkan bahwa algoritma k-means mampu menghasilkan pemilihan dan pembagian kelas unggulan sesuai nilai kemampuan siswa di SMP Pelita Bandung.

Kata kunci: *Data Mining*, Kelas Unggulan, *K-Means Clustering*

I. PENDAHULUAN

Pendidikan saat ini menjadi kebutuhan manusia. Pendidikan merupakan suatu cara pembelajaran untuk meningkatkan pengetahuan, sikap dan keterampilan seseorang. Sekolah menjadi sarana berlangsungnya pendidikan secara langsung, dengan bertemunya siswa dan guru dapat terjadi kegiatan belajar mengajar yang lebih

efektif dan efisien (Amirulloh, 2019 [1]). Pentingnya pendidikan dalam perannya mencerdaskan kehidupan bangsa. Oleh karena itu, setiap bangsa hendaknya memiliki pendidikan yang baik dan berkualitas (Hadinata, 2019 [2]).

Di Indonesia Sekolah Menengah Pertama (SMP) merupakan pendidikan formal pada jenjang pendidikan dasar. Waktu yang ditempuh sekolah menengah pertama 3 tahun mulai dari kelas 7 sampai dengan kelas 9. Pendidikan dan pembelajaran di tingkat SMP memberikan pondasi dalam

^{*)} novi.eung123@gmail.com

menyiapkan generasi agar menjadi manusia yang mampu menghadapi era yang semakin berkembang.

Penilaian prestasi siswa sangat dibutuhkan karena setiap penentuan prestasi siswa didapat dari hasil proses prestasi setiap siswa di sekolah, baik dalam kegiatan pembelajaran di kelas, maupun dalam kegiatan ekstrakurikuler (Gunawan, 2019 [3]). Penilaian merupakan rangkaian kegiatan untuk memperoleh, menganalisis dan menafsirkan data tentang proses hasil belajar siswa-siswi yang dilakukan secara sistematis dan berkesinambungan sehingga dapat dijadikan dasar informasi pengambilan keputusan yang bermakna (Winaryati, 2018 [4]).

Dalam sekolah umum terdapat sejumlah siswa yang belajar di sejumlah kelas, untuk kualitas seorang siswa diperlukan adanya suatu kelas yang berisikan siswa terbaik yang dinamakan sebagai kelas unggulan. Kelas Unggulan merupakan kelas siswa-siswi yang unggul dengan kecerdasan diatas rata-rata yang dikelompokkan secara khusus dengan maksud untuk membimbing siswa dalam mengembangkan kecerdasan, kemampuan, keterampilan, dan sikap yang terbaik (Ramadani, 2017 [5]).

SMP Pelita Bandung adalah sekolah menengah pertama yang beralamat di Jalan Cikutra Kota Bandung. Sekolah ini memiliki

proses pembelajaran yang baik. Setiap tahun minat pendaftaran di sekolah ini meningkat sehingga dijadikan sekolah swasta favorit di Kota Bandung. Setiap tahunnya pendaftar semakin bertambah banyak sehingga meningkatnya jumlah data siswa. Kelas unggulan yang telah dibentuk dirasa masih kurang efektif sehingga kesulitan untuk menentukan nilai siswa-siswi untuk dimasukan ke kelas unggulan sesuai dengan nilai siswa.

Salah satu untuk mengatasi masalah yang ada adalah dengan menggunakan penerapan *Data Mining* yang bisa digunakan untuk pengolahan data menjadi sumber informasi strategis dan lebih mudah bagi sekolah menyeleksi siswa masuk kelas unggulan maka peneliti ini menggunakan metode *Algoritma K-Means Clustering*. *Algoritma K-Means* adalah salah satu metode pengelompokan data nonhirarki yang mempartisi data. Metode ini mempartisi data dengan karakteristik yang sama ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok data yang lain (Rahmawati, 2020 [6]).

1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas, maka identifikasi masalah yang adadiantaranya:

1. Bagaimana menentukan kelas unggulan menggunakan metode *K-Means*

Clustering di SMP Pelita Bandung?

2. Bagaimana memanfaatkan teknik data mining dengan menggunakan algoritma *K-Means* untuk menyeleksi kelas unggulan di SMP Pelita Bandung?

1.3 Tujuan Penelitian

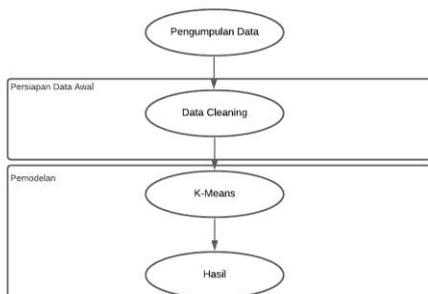
Tujuan dari penulis membuat penelitian ini adalah sebagai berikut:

1. Untuk menentukan cluster siswa dalam pengelompokan kelas unggulan berdasarkan data nilai siswa di SMP Pelita Bandung.
2. Memberikan langkah strategis bagi SMP Pelita Bandung dalam menyeleksi kelas unggulan dengan metode *K-Means Clustering*.

II. METODOLOGI

2.1 Tahapan Penelitian

Tahapan penelitian bertujuan untuk menyelesaikan masalah yang ada secara terstruktur. Berikut adalah tahapan yang dilakukan peneliti ini seperti pada Gambar 1.



Gambar 1.
Desain Penelitian

1. Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data. Pengumpulan data merupakan salah satu tahapan utama untuk mendapatkan data. Data yang digunakan peneliti adalah data public yang diambil langsung dengan cara riset kepada bagian kesiswaan sekolah. Teknik pengumpulan data yang benar akan menghasilkan data yang memiliki integritas tinggi.

2. Data Cleaning

Dalam penelitian ini sebelum proses data mining dilakukan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses Data Cleaning merupakan suatu proses mendeteksi dan memperbaiki atau menghapus data set, tabel, dan database yang tidak akurat, tidak tepat, tidak lengkap, dan tidak benar yang kemudian *dirty* data tersebut akan diganti, dimodifikasi atau dihapus. Proses data cleaning ini cukup penting untuk mencegah terjadinya duplikat data, ambigu pada data dan memperbaiki kesalahan pada data.

3. K-Means

Pada tahap ini metode yang digunakan adalah K-Means. K-Means merupakan algoritma yang digunakan peneliti karena K-Means adalah algoritma yang mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang

memiliki karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Aplikasi yang digunakan pada penelitian ini adalah *Rapidminer*.

4. Hasil

Hasil dari pengujian menggunakan *rapidminer* yang digunakan untuk pemodelan dataset pada Algoritma *K-Means* menghasilkan beberapa kesimpulan yang digunakan sebagai hasil penelitian.

2.2 Instrumen Penelitian

Instrumen penelitian merupakan suatu alat yang memenuhi persyaratan akademis sehingga dapat digunakan sebagai alat untuk mengukur suatu objek ukur atau mengumpulkan data mengenai suatu variabel (Sappaile, 2007 [7]). Dengan instrumen yang sesuai dan tepat akan menghasilkan data yang benar dengan kesimpulan yang sesuai keadaannya. Penelitian yang digunakan pada penelitian ini antara lain:

1. Perangkat Lunak (*Software*)

Software Rapidminer versi 9.4 menjadi perangkat lunak yang digunakan pada penelitian ini.

2. *Dataset*

Dataset pada penelitian ini menggunakan dataset data nilai siswa kelas VIII SMP Pelita Bandung Tahun Ajaran 2020-2021 sebanyak 119 data siswa yang telah melalui proses *cleaning* terdiri dari atribut.

2.3 Metode Pengumpulan Data

Penelitian ini menggunakan data *public* yang di ambil langsung dari SMP Pelita Bandung dengan cara melakukan riset kepada bagian kesiswaan SMP Pelita Bandung.

A. Sampel Data

Dalam penelitian ini sampel data yang digunakan adalah data siswa kelas VIII sebanyak 119 data siswa terdiri dari atribut NIS, Nama Siswa, Nilai mata pelajaran Bahasa Indonesia, Matematika, IPA, dan Bahasa Inggris.

Tabel 1.
Atribut Data

NIS	Nomor Induk Siswa SMP Pelita Bandung
Nama Siswa	Nama Siswa SMP Pelita Bandung
Nilai	Nilai yang diambil dari raport siswa

2.4 Metode Pengumpulan Data

A. Pengolahan Data Awal

Pada tahap metode pengumpulan data merupakan tahap untuk memastikan data siswa yang dipilih telah layak untuk dilakukan proses pengolahan data *cleaning*.

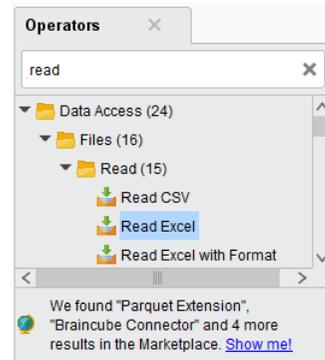
Data *cleaning* merupakan proses data yang tidak konsisten atau data tidak relevan. Data-data yang tidak relevan itu juga lebih baik

dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya [8]. Maka *dataset* ini hanya diambil 119 data dari 202 data, karna adanya data yang tidak bisa diolah dan hanya mengambil 20 besar dari setiap kelas untuk diolah. Maka pada proses *cleaning* ini dilakukan penghapusan untuk menghilangkan *missing value* pada *dataset*.

B. Permodelan

Tahap pemodelan ini dilakukan menggunakan software rapidminer 9.4 dengan langkah sebagai berikut:

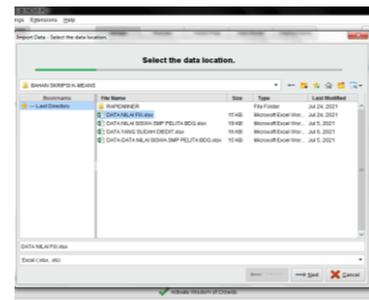
1. Buka *software rapidminer*.
2. Pilih menu *file* klik *new process*.
3. Di bagian bawah terdapat *panel operator* seperti yang terlihat pada Gambar 2 kemudian ketik *read excel*, lalu *double klik* atau *drag* pada *read excel* agar *operator* tersebut muncul di *panel process* dan bisa digunakan untuk memasukkan data *excel* yang melalui tahap data *cleaning*.



Gambar 2.

Pilih Read Excel

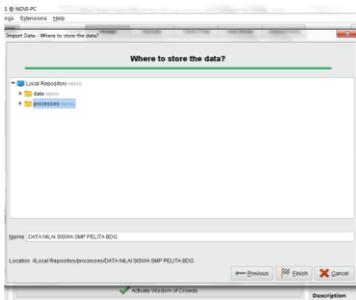
4. Di bagian samping terdapat Parameters kemudian klik *import configuration wizard* agar bisa memilih file *database* yang akan diproses di *rapidminer*.



Gambar 3.

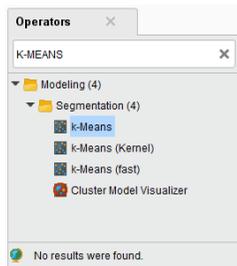
Memilih Dataset

5. Setelah proses pada gambar 4 kemudian simpan pada *process*.



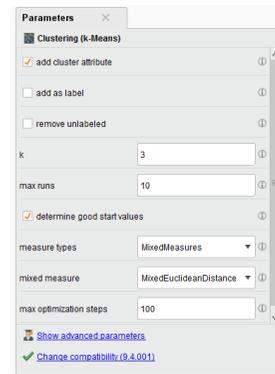
Gambar 4.
Simpan Dataset

6. Selanjutnya memasukan model untuk digunakan proses *clustering* data mining dengan cara masuk ke *operator* kemudian ketik *K-Means*.



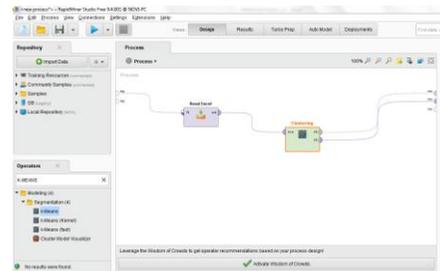
Gambar 5.
Pilih Pemodelan

7. Pada proses *clustering* tentukan jumlah klasternya menjadi $K=3$ dan pilih *mixed measure* menjadi *mixed euclidean distance* seperti Gambar 6.



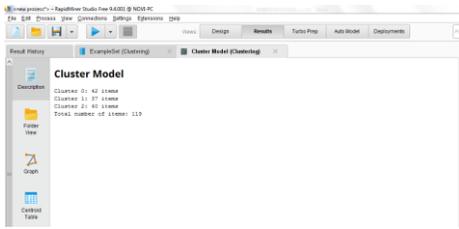
Gambar 6.
Menentukan Jumlah K

8. Selanjutnya hubungkan *conector* masing-masing proses pada *main process* setelah itu klik *button run*.



Gambar 7.
Menggabungkan Conector

9. Pada Gambar 8 hasil pengujian *dataset* yang berjumlah 119 data menggunakan *software rapidminer 9.4* terbentuk 3 *cluster*. Pada *cluster 0* (*cluster* pertama) terdapat 42 *items*, pada *cluster 1* (*cluster* kedua) terdapat 37 *items*, dan pada *cluster 2* (*cluster* ketiga) terdapat 40 *items*.



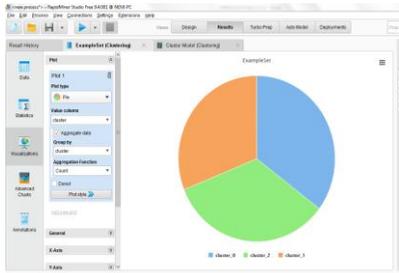
Gambar 8.
Hasil Cluster

10. Dari 3 cluster yang telah terbentuk kemudian centroid akhir untuk setiap cluster, adalah cluster 0 = 181907186, 80.619, 73.667, 84.476 cluster 1 = 181907034, 84.838, 74.973, 81.324, 88.541 cluster 2 = 181907104, 81.700, 73.975, 79.550, 87.400 seperti pada Gambar 9.

Observation	cluster_0	cluster_1	cluster_2
AG	181907186	181907034	181907104
BAHASA INDONESIA	80.619	84.838	81.700
MATEMATIKA	73.667	74.973	73.975
IPA	79.238	81.324	79.550
BAHASA INGGRIS	84.476	88.541	87.400

Gambar 9.
Tabel Centroid

11. Berikut *chart* dari setiap jumlah *cluster* seperti Gambar 10.

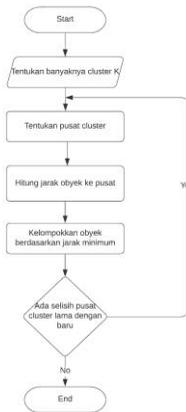


Gambar 10.
Pie Cluster

III. ANALISIS DAN PEMBAHASAN

3.1 Hasil Perhitungan

A. Flowchart Pengolahan Data



Gambar 11.

Flowchart K-Means

Pada Flowchart diatas untuk melakukan perhitungan manual, maka penjelasannya sebagai berikut :

1. Pertama siapkan data agar data bisa dihitung/diolah.
2. Tentukan banyaknya *cluster* K sesuai yang dibutuhkan.
3. Setelah jumlah *cluster* K telah ditentukan maka hitung jarak ke pusat agar mengetahui hasil data yang akan dikelompokkan pada setiap *cluster*.
4. Jarak objek ke pusat telah diketahui hasilnya, selanjutnya dapat dihitung jarak minimum data untuk dikelompokkan.
5. Data yang sudah dikelompokkan berdasarkan jarak minimum dan *cluster*, maka akan diketahui pusat *cluster* baru yang akan dihitung kembali

6. Pusat cluster baru yang sudah diketahui maka dihitung kembali, jika perhitungan antara cluster baru dan cluster lama ada selisih (berubah) data maka perhitungan diulang kembali, tetapi jika perhitungan antara cluster baru dan cluster lama sudah tidak ada selisih (tidak berubah) perhitungan bisa berhenti.

B. Perhitungan K-Means

Dalam menentukan banyaknya *cluster* k dapat dilakukan dengan beberapa pertimbangan seperti pertimbangan terioritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*. Penetapan jumlah *cluster* k pada penelitian ini yaitu berjumlah 3 *cluster*.

4.2 Hasil Penelitian dan Pembahasan

Hasil penelitian yang telah di uji antara *Microsoft Excel* dan *Software Rapidminer* yaitu mendapatkan hasil yang sama. Pada perhitungan data di *Microsoft Excel* menemukan hasil yang tidak berubah yaitu pada iterasi 4 dengan 3 cluster dan beberapa atribut seperti Nama Siswa, NIS, Nilai (B.Indonesia, Matematika, IPA, B.Ingggris).

Tabel 2.
Centroid Baru

Titik Pusat Awal	NIS	B.Indonesia
Cluster 0	181907185.4	80.5116279
Cluster 1	181907034.6	84.8378378
Cluster 2	181907103.8	81.8461538

Matematika	IPA	B.Ingggris
73.58139535	78.20930233	84.53488372
74.97297297	81.31434324	88.54054054
74.07692308	79.61538462	87.41025641

Untuk memastikan hasil perhitungan manual *Microsoft Excel* dan perhitungan *Software Rapidminer 9.4* dengan hasil yang sama.



Gambar 12.
Tabel Centroid

Berdasarkan dari hasil perhitungan diatas yang telah dibuat, dari 119 data siswa di bagi menjadi kedalam 3 *cluster* berdasarkan NIS, Nilai B.Indonesia, Nilai Matematika, Nilai IPA, dan Nilai B.Ingggris. Pada *cluster* 0 yang telah dihitung berdasarkan *Microsoft excel* dan *Rapidminer* menghasilkan jumlah 42 siswa. Pada *cluster* 1 yang telah dihitung berdasarkan *Microsoft excel* dan *Rapidminer*

menghasilkan jumlah 37 siswa. Dan pada *cluster* 2 yang telah dihitung berdasarkan *Microsoft excel* dan *Rapidminer* menghasilkan jumlah 40 siswa.

Dari 3 *cluster* yang terbentuk, yaitu *cluster* 0 adalah *cluster* dengan nilai rata-rata terendah akan dimasukkan kedalam kelas C sebanyak 42 siswa, pada *cluster* 1 dengan nilai rata-rata sedang akan dimasukkan kedalam kelas B sebanyak 37 siswa, sedangkan pada *cluster* 3 dengan nilai rata-rata tertinggi akan dimasukkan kedalam kelas A sebanyak 40 siswa.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan pada perhitungan menggunakan Rapidminer dan Microsoft Excel maka dapat disimpulkan bahwa SMP Pelita Bandung membutuhkan 3 cluster. Setelah peneliti melakukan eksperimen dapat menghasilkan 3 cluster yaitu *cluster* 0 adalah *cluster* dengan nilai rata-rata terendah akan dimasukkan kedalam kelas C sebanyak 42 siswa, pada *cluster* 1 dengan nilai rata-rata sedang akan dimasukkan kedalam kelas B sebanyak 37 siswa, sedangkan pada *cluster* 3 dengan nilai rata-rata tertinggi akan dimasukkan kedalam kelas A sebanyak 40 siswa. Hasil dari penelitian ini menunjukkan bahwa algoritma k-means mampu menghasilkan pemilihan dan

pembagian kelas unggulan sesuai nilai kemampuan siswa.

Dengan adanya penerapan data mining algoritma K-Means diharapkan mampu memberikan solusi dalam menentukan kelas unggulan di SMP Pelita Bandung.

DAFTAR PUSTAKA

- [1] Amirulloh, I. (2019). Pemetaan Kelompok Kerja Siswa Dengan Metode Clustering K-Means Dan Algoritma Greedy. *J. Inform. dan Rekayasa Perangkat Lunak*, Vol. 1(294–98). DOI: 10.36499/jinrpl.v1i2.2953.
- [2] Hadinata, N.S, Fitriyani, F. (2019). Website Pengolahan Nilai Siswa menggunakan Konsep Model - View - Controller pada LIA Pamulang. *J. Ilm. Inform. Glob.*, Vol. 10(1): 38–43. DOI: 10.36982/jig.v10i1.746.
- [3] Gunawan, I., Sumarno, S., Tambunan, H.S. (2019). Penggunaan Algoritma Sorting Bubble Sort Untuk Penentuan Nilai Prestasi Siswa. *Sistemasi*, Vol. 8(2): 296. DOI: 10.32520/stmsi.v8i2.493.
- [4] Winaryati, E. (2018). Penilaian Kompetensi Siswa Abad 21. *Pros. Semin. Nas. Int.*, vol. 6 (1): 6–19.
- [5] Ramadani, S.F., Ekojono, E., Santoso, N. (2017). Sistem Pendukung Keputusan Seleksi Siswa Kelas Unggulan Di Smp

- Negeri 7 Malang. *J. Inform. Polinema*, Vol. 3(3): 27 DOI: 10.33795/jip.v3i3.30.
- [6] Rahmawati, Arifin, T. (2020). Penerapan Algoritma K-Means Untuk Pengelompokan Siswa Lolos SNMPTN Di SMAN 8 Bandung. *J. Responsif*, Vol. 2(2): 184–190.
- [7] Sappaile, B.I. (2007). Konsep Instrumen Penelitian Pendidikan. *J. Pendidik. dan Kebud.*, Vol. 13(66): 379-391 DOI: 10.24832/jpnk.v13i66.356.
- [8] Purnia, D.S., Warnilah, A.I. (2017). Implementasi Data Mining Pada Penjualan kacangata Dengan Menggunakan Algoritma Apriori. *Indones. J. Comput. Inf. Technol.*, Vol. 2(2): 31–39. DOI: <https://doi.org/10.31294/ijcit.v2i2.2776>.