
Penerapan *K-Means Clustering* Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode *Elbow*

Adella Putri Riani¹, Apriade Voutama², Taufik Ridwan³

^{1,2,3} Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Email: ¹adella.putri19001@student.unsika.ac.id, ²apriade.voutama@staff.unsika.ac.id, ³taufik.ridwan@cs.unsika.ac.id

Email Penulis Korespondensi: adella.putri19001@student.unsika.ac.id

Article History:

Received Nov 07th, 2022

Revised Dec 20th, 2022

Accepted Jan 03th, 2023

Abstrak

Pendidikan merupakan salah satu faktor keberhasilan teknologi informasi salah satunya yaitu dalam hal memajukan mutu sumber daya manusia. Salah satu cara untuk mencapainya adalah melalui pemerataan pendidikan di sekolah. Hal tersebut dapat dilakukan dengan menganalisis capaian hasil belajar setiap peserta didik. PKBM Sanggar Puri merupakan lembaga pendidikan non formal dengan nilai nilai rata-rata dari tahun ke tahun naik turun sehingga perlu dilakukan pengelompokan nilai guna membantu mengetahui kemampuan para peserta didik. Penelitian ini menggunakan algoritma *K-Means* dan metode *Elbow*. Hasil dari penelitian ini adalah mengelompokan peserta didik menjadi 3 *cluster* yaitu *cluster 0* dengan nilai rata-rata cenderung rendah yang memiliki jumlah anggota *cluster* yaitu 57 peserta didik. Kemudian *cluster 2* yaitu kelompok rata-rata nilai sedang dengan jumlah anggota *cluster* yaitu 42 peserta didik. Lalu *cluster 1* merupakan kelompok nilai dengan rata-rata tinggi yang hanya berjumlah 1 peserta didik.

Kata Kunci : *Elbow, Cluster, K-Means*

Abstract

Education is one of the success factors of information technology, one of which is in terms of advancing the quality of human resources. One way to achieve this is through equal distribution of education in schools. This can be done by analyzing the learning outcomes of each student. PKBM Sanggar Puri is a non-formal educational institution with an average value from year to year up and down so it is necessary to group values to help determine the abilities of the students. This study uses the K-Means algorithm and the Elbow method. The results of this study are to group students into 3 clusters, namely cluster 0 with an average value that tends to be low which has a number of cluster members, namely 57 students. Then cluster 2 is the average group with a moderate value with the number of cluster members, namely 42 students. Then cluster 1 is a group of values with a high average which only amounts to 1 student.

Keyword : *Cluster, Elbow, K-Means*

1. PENDAHULUAN

Perkembangan teknologi informasi saat ini telah membuktikan suatu kemajuan yang tidak dapat dihindari dan turut memberikan imbas positif pada berbagai bidang seperti ekonomi, kesehatan, seni, pendidikan, dan lain-lain. Namun perkembangan tersebut juga perlu dibarengi dengan peningkatan mutu SDM dalam faktor menentukan keberhasilan perkembangan teknologi informasi. Teknologi yang berkembang pesat menuntut penggunaannya, mau tidak mau menggunakan teknologi dalam memaksimalkan pekerjaan, salah satunya adalah pendidikan [1]. Pendidikan menjadi salah satu aspek esensial dalam hal peningkatan mutu SDM. Kualitas pendidikan yang baik tentunya harus diimbangi dengan proses belajar mengajar yang merata. Hal tersebut dapat dilakukan dengan menganalisa capaian hasil belajar untuk menentukan kompetensi peserta didik dalam kegiatan belajar mengajar (KBM) di kelas. Hasil analisa tadi dapat membantu tenaga pendidik dalam mengoptimalkan metode pembelajaran untuk pemerataan KBM di kelas.

PKBM Sanggar Puri mengadakan program pendidikan kesetaraan bagi masyarakat yang ingin mendapatkan ijazah. Setiap lembaga pendidikan pasti menginginkan lulusan yang unggul dalam berbagai disiplin ilmu. Prestasi merupakan hasil kemampuan belajar peserta didik dari proses KBM yang telah dilaksanakan selama pembelajaran di kelas [2]. Proses

belajar mengajar di PKBM Sanggar Puri tidak jauh berbeda dengan sekolah formal. Namun nilai rata-rata mata pelajaran Ujian Nasional (UN) kelas paket C di PKBM Sanggar Puri dari semester ke semester naik turun. Oleh karena itu maka perlu dilakukan pengelompokan nilai guna membantu mengetahui kemampuan para peserta didik sehingga dapat dengan mudah diketahui dan diidentifikasi mana saja peserta didik yang masih kurang dalam pembelajaran dan dapat mempersiapkan model pembelajaran yang lebih tepat di semester selanjutnya.

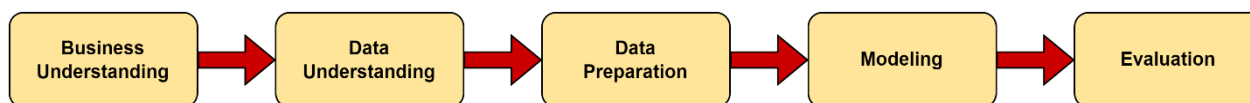
Clustering adalah metode pengelompokan data yang membagi menjadi banyak *cluster* (kelompok) dengan kemiripan maksimal antara data pada 1 *cluster* dan memiliki kemiripan yang minimal dengan *cluster* lain [3]. Teknik *Clustering* dapat mengatasi permasalahan di atas yaitu dengan melakukan pengelompokan hasil belajar peserta didik berdasarkan nilai. Penelitian ini menggunakan metode *K-Means*. Metode *K-Means* merupakan teknik data *clustering* yang membagi data menjadi beberapa kelompok-kelompok sesuai dengan fitur data tersebut [4]. Algoritma *K-Means* memiliki beberapa kelebihan yaitu mampu mengagregasi data dalam objek yang besar dengan proses yang singkat dan efisien [5]. Namun algoritma *K-Means* juga memiliki kekurangan yaitu hasil *cluster* bergantung pada nilai *c* atau jumlah *cluster* [6]. Maka dari itu, peneliti dalam penelitian ini menggunakan metode *Elbow* untuk menentukan nilai jumlah *cluster* (*c*) terbaik sehingga menghasilkan *cluster* yang optimal.

Beberapa penelitian sudah pernah dilakukan menggunakan metode *K-Means* diantaranya adalah penelitian yang dilakukan Vickry Ramadhan dan Apriade Voutama yang berjudul “*Clustering Menggunakan Algoritma K-Means Pada Penyakit ISPA di Puskesmas Kabupaten Karawang*” dimana dari 50 puskesmas jumlah yang diuji terdapat 30 puskesmas di *cluster* 1 yaitu puskesmas dengan tingkat kasus ISPA rendah, 19 puskesmas di *cluster* 2 dengan tingkat kasus ISPA sedang dan 1 puskesmas di *cluster* 3 dengan tingkat kasus ISPA tinggi [7]. Teti Purwanti, William Ramdhan, Santoso dalam penelitian yang berjudul “*Penerapan Metode Clusterisasi K-Means untuk Strategi Promosi pada SMK Tamansiswa Sukadama*” menghasilkan 2 *cluster* dimana *cluster* 1 terdapat 14 wilayah yang berpotensi untuk promosi, dan pada *cluster* 2 terdapat 4 wilayah yang tidak berpotensi untuk dilakukan promosi [8].

Percobaan yang dilakukan Ninik yang berjudul “*Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional dalam menentukan jumlah cluster terbaik yaitu dengan membandingkan selisih nilai Sum of square Error (SSE) terbesar dengan jumlah data yang berbeda sebanyak 5 kali uji coba dan mendapatkan jumlah cluster terbaik yaitu 3 cluster* [9]. Ridho Ananda dan Achmad Zaki Yamani dalam penelitiannya yang berjudul “*Penentuan Centroid Awal K-Means pada proses Clustering Data Evaluasi Pengajaran Dosen*” menghasilkan kesimpulan bahwa hasil *K-Means Clustering* dengan kriteria *Elbow* akan menghasilkan kualitas *cluster* yang lebih baik daripada *K-Means* tanpa kriteria apapun dengan peluang sebesar 0.778 [10]. Adapun tujuan dari penelitian ini adalah melakukan pengelompokan nilai peserta didik paket C PKBM Sanggar Puri sehingga dapat digunakan sebagai bahan pertimbangan dalam menerapkan model pembelajaran di PKBM Sanggar Puri. Nilai yang digunakan adalah nilai rata-rata mata pelajaran UN dari semester 1 hingga semester 5.

2. METODOLOGI PENELITIAN

Pendekatan dalam penelitian ini adalah metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM memiliki lima tahapan dan kerangka kerja sesuai pada gambar 1 sebagai berikut:



Gambar 1. CRISP-DM

2.1 Business Understanding

Tahapan ini berfokus pada penentuan masalah bisnis dan pemahaman lingkungan penelitian. Masalah bisnis yang akan dipecahkan harus diuraikan secara jelas dan terukur. Tahap pemahaman bisnis dapat dilakukan dengan memahami tujuan jangka panjang dari sebuah bisnis [11]. Setelah memahami tujuan jangka panjang bisnis langkah selanjutnya yaitu memahami masalah bisnis yang terjadi serta aspek-aspek apa saja yang mempengaruhi masalah tersebut. Setelah mendapatkan pemahaman mengenai tujuan dan masalah bisnis yang terjadi tahap selanjutnya adalah menentukan bagaimana cara atau penyelesaian untuk mengatasi masalah tersebut.

2.2 Data Understanding

Tahap ini merupakan tahap mengumpulkan data awal, mengeksplorasi data, dan mengecek kualitas data. Tahapan ini bertujuan untuk mendapatkan gambaran awal mengenai data. Data-data tersebut dapat diambil dari database, data warehouse, internet, dan lain-lain kemudian diolah menjadi informasi [12].

2.3 Data Preparation

Tahap ini merupakan tahap dimana data mentah diubah menjadi bentuk yang mudah dimengerti. Tahap ini melibatkan proses *data integration* yaitu proses menggabungkan data dari beberapa sumber, *data cleaning* yaitu langkah pembersihan data termasuk menghapus atau mengganti nilai null [13], serta *data selection* yaitu tahap seleksi data sesuai dengan fokus penelitian yang akan dilakukan serta. Proses penyesuaian data sangat penting agar sesuai dengan kebutuhan penelitian [14].

2.4 Modelling

Pemodelan merupakan tahap pemilihan model dan penerapan pemodelan dengan algoritma data mining. Tahap ini bertujuan untuk mengoptimalkan hasil penelitian. Pada penelitian ini terdapat 2 pemodelan yang dilakukan.

2.4.1 Metode Elbow

Metode *Elbow* merupakan teknik yang digunakan untuk menentukan jumlah *cluster* (c) optimal dengan cara menghitung nilai *Sum of Square Error* (SSE) dari masing-masing *cluster*. Semakin besar selisih nilai SSE dengan *cluster* selanjutnya sehingga membentuk titik sudut siku, maka semakin baik jumlah *cluster* [15]. Nilai c pada kombinasi *elbow* dan *K-Means* merupakan plot hubungan *cluster* dengan *error* yang direduksi [16]. Peningkatan nilai c menyebabkan grafik menurun secara perlahan hingga nilai k stabil [16]. SSE banyak digunakan untuk mendapatkan jumlah c yang optimal. Metode ini menguji nilai SSE dengan jumlah *cluster* yang berbeda-beda kemudian mencari nilai SSE dengan selisih terbesar atau yang membentuk sudut paling siku pada grafik *elbow* dalam menentukan jumlah *cluster* terbaik [17]. Rumus SSE dapat dilihat pada persamaan 1.

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (1)$$

2.4.2 K-Means Clustering

Algoritma *K-Means* merupakan teknik analisis data yang menggunakan sistem partisi untuk melakukan proses pengelompokan data [2]. Suatu data dikelompokkan ke dalam satu *cluster* berdasarkan kemiripan atribut yang dimiliki. Kemiripan ini bisa diketahui dengan mengukur jarak setiap data dengan pusat *cluster* (*centroid*) [13]. Metode *K-Means Clustering* merupakan salah satu metode *Clustering* untuk mengelompokkan data yang memiliki jumlah data besar dan proses yang cepat dan efisien [8]. Berikut adalah langkah perhitungan algoritma *K-Means*

- Tentukan berapa banyak jumlah *cluster* (c) atau kelompok
- Tentukan secara acak pusat *cluster* awal (*centroid*)
- Ukur jarak setiap data dengan pusat *cluster* (*centroid*) yaitu dengan rumus Euclidean Distance pada persamaan 2.

$$D_{i,j} = \sqrt{(x_{1p} - x_{pq})^2 + (x_{2p} - x_{2q})^2 + \dots + (x_{rp} - x_{rq})^2} \quad (2)$$

Keterangan :

$D_{(p,q)}$ = jarak data ke- p dengan pusat *cluster* q

$X_{(r,p)}$ = data ke- p pada atribut data ke- r

$X_{(r,q)}$ = titik pusat ke- q pada atribut

- Kelompokkan setiap data ke dalam *cluster* berdasarkan jarak minimum
- Lakukan proses iterasi dengan menentukan pusat *cluster* (*centroid*) baru dengan rumus pada persamaan 3.

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

Keterangan :

v = *centroid* pada *cluster*

$x_{(p)}$ = objek ke- p

n = banyaknya objek/jumlah

- Ulangi langkah 3 hingga tidak terdapat perubahan *cluster* pada setiap data dari proses iterasi sebelumnya.

2.5 Evaluation

Tahap evaluasi bertujuan untuk mengukur kinerja dan akurasi dari pemodelan yang telah diperoleh sesuai dengan target yang ingin dicapai pada tahap pertama. Pada penelitian ini dalam mengukur kinerja dan akurasi model yang telah dibuat menggunakan *Davies Bouldin Index*.

2.5.1 Davies Bouldin Index

Davies Bouldin Index merupakan teknik evaluasi *cluster* berdasarkan kesamaan ukuran *cluster* dan perbedaan ukuran *cluster* dalam distribusi *cluster* [18]. *Davies Bouldin Index* bertujuan untuk mengukur performa sebuah model dalam

menentukan jumlah *cluster* yang terbaik. Semakin kecil nilai DBI maka semakin optimal jumlah *cluster* [16]. Rumus DBI dapat dilihat pada persamaan 4.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \tag{4}$$

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Objek pada penelitian ini yaitu PKBM Sanggar Puri. Masalah bisnis pada PKBM Sanggar Puri adalah rata-rata nilai pada mata pelajaran yang diujikan dalam ujian nasional (UN) peserta didik paket C naik turun. Dari permasalahan tersebut maka tujuan dari penelitian ini adalah bagaimana cara mengetahui kemampuan akademik untuk meningkatkan kompetensi peserta didik paket C PKBM Sanggar Puri khususnya pada mata pelajaran UN. Oleh sebab itu dalam penelitian ini akan dilakukan penggalan data nilai rata-rata peserta didik paket C dari semester 1 sampai semester 5 dengan metode pengelompokan menggunakan algoritma *K-Means*.

3.2 Data Understanding

Berdasarkan pemahaman bisnis dan masalah yang terjadi, data yang diperlukan yakni data nilai peserta didik paket C kelas 12 jurusan IPA dan IPS dari tahun 2019-2022 dengan jumlah peserta yaitu 100 peserta didik. Tahap ini merupakan tahap untuk mengeksplorasi dan mempelajari data yang didapatkan. Data yang telah didapatkan berupa nilai akhir 14 mata pelajaran paket C di PKBM Sanggar Puri mulai dari semester 1 sampai dengan semester 5.

3.3 Data Preparation

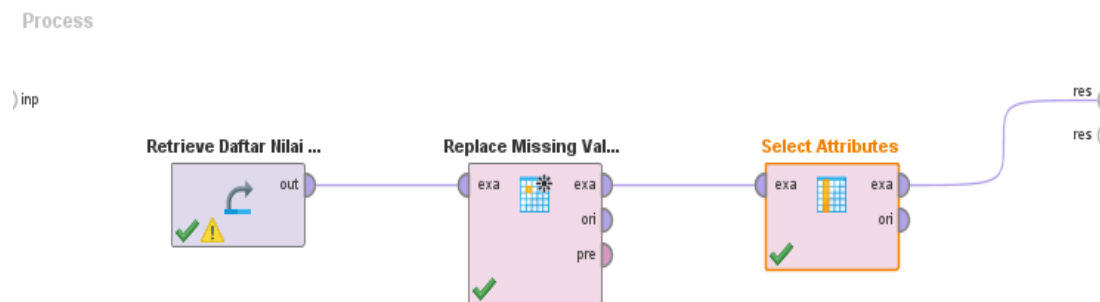
Tahap selanjutnya yaitu melakukan proses persiapan data yang terdiri *data integration*, *data cleaning*, dan *data selection*. Tahap pertama yaitu tahap penggabungan data atau *data integration*. Proses ini menggabungkan nilai akhir 14 mata pelajaran dari semester 1-5 menjadi nilai rata-rata tiap semester. Selanjutnya setelah tahap penggabungan data tahap berikutnya yaitu pembersihan data. Proses *data cleaning* ini menggunakan bantuan aplikasi *Rapid Miner*. Pada tahap ini tidak terdapat nilai null atau nilai yang kosong sehingga dapat dilanjutkan ke tahap berikutnya yaitu *data integration*. Selanjutnya pada tahap *data selection* atau pemilihan data, dari total 20 atribut hanya 6 atribut saja yang akan digunakan. Atribut tersebut antara lain mata pelajaran wajib bahasa Indonesia, matematika, bahasa Inggris. Kemudian mata pelajaran peminatan IPA yaitu biologi, fisika, kimia dan geografi, ekonomi, sosiologi untuk jurusan IPS seperti yang dapat dilihat pada Tabel 1.

Tabel 1. Nilai Rata-Rata Mata Pelajaran UN Paket C PKBM Sanggar Puri

| No | Bahasa Indonesia | Matematika | Bahasa Inggris | Biologi / Biografi | Fisika / Ekonomi | Kimia / Sosiologi |
|-----|------------------|------------|----------------|--------------------|------------------|-------------------|
| 1 | 81 | 80 | 80 | 81 | 82 | 81 |
| 2 | 81 | 81 | 81 | 82 | 82 | 81 |
| 3 | 82 | 81 | 82 | 82 | 82 | 82 |
| ... | | | | | | |
| 24 | 85 | 83 | 84 | 83 | 84 | 84 |
| 25 | 86 | 85 | 84 | 83 | 85 | 85 |
| 26 | 86 | 85 | 85 | 85 | 87 | 86 |
| ... | | | | | | |
| 58 | 92 | 92 | 90 | 93 | 91 | 92 |
| 59 | 82 | 80 | 80 | 81 | 82 | 81 |
| 60 | 83 | 80 | 80 | 83 | 80 | 81 |
| ... | | | | | | |
| 86 | 81 | 82 | 81 | 81 | 81 | 81 |
| 87 | 83 | 83 | 82 | 81 | 81 | 82 |
| 88 | 82 | 83 | 82 | 80 | 81 | 82 |

| | | | | | | |
|-----|----|----|----|----|----|----|
| ... | | | | | | |
| 98 | 87 | 85 | 87 | 81 | 82 | 84 |
| 99 | 87 | 86 | 88 | 79 | 84 | 85 |
| 100 | 87 | 86 | 86 | 83 | 81 | 85 |

Proses Persiapan Data dengan Aplikasi Rapid Miner, seperti yang dapat dilihat pada gambar 2.



Gambar 2. Proses Persiapan Data dengan Aplikasi *Rapid Miner*

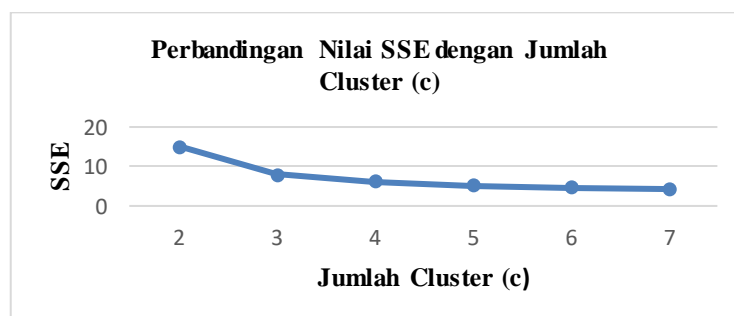
3.4 Modelling

Setelah data siap dipakai langkah selanjutnya yaitu melakukan pemodelan. Pada penelitian ini dalam menentukan jumlah *cluster* (*c*) yang optimal adalah menggunakan metode *elbow*. Dengan metode *elbow* jumlah *cluster* dikatakan optimal apabila grafik perbandingan *Sum of Square Error* (SSE) dengan jumlah *cluster* membentuk siku (*elbow*). Artinya, selisih nilai SSE dengan *cluster* sebelumnya berubah drastis namun selisih nilai SSE dengan *cluster* setelahnya tidak berubah drastis. Di bawah ini tabel 2 merupakan nilai SSE dengan jumlah *cluster* mulai dari *c*=2 sampai *c*=7.

Tabel 2. Perbandingan Nilai SSE dengan Jumlah *Cluster*

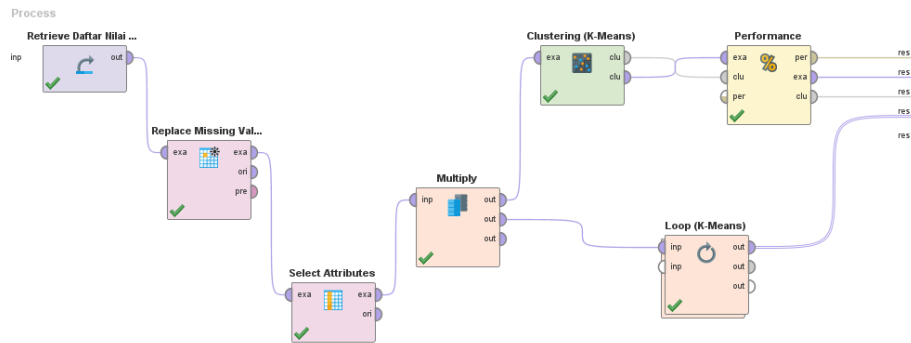
| No | Jumlah <i>Cluster</i> (<i>c</i>) | <i>Sum of square Error</i> (SSE) |
|----|------------------------------------|----------------------------------|
| 1 | 2 | 15,181 |
| 2 | 3 | 7,855 |
| 3 | 4 | 6,152 |
| 4 | 5 | 5,189 |
| 5 | 6 | 4,561 |
| 6 | 7 | 4,278 |

Di bawah ini gambar 3 merupakan Grafik Perbandingan Nilai SSE dengan Jumlah *Cluster* (*c*).



Gambar 3. Grafik Perbandingan Nilai SSE dengan Jumlah *Cluster* (*c*)

Berdasarkan gambar 3 dapat dilihat garis yang paling membentuk siku (*elbow*) atau memiliki sudut yang paling lancip yaitu terletak saat jumlah *cluster* $c = 3$. Artinya $c = 3$ merupakan jumlah *cluster* terbaik dengan nilai SSE yaitu 7,855. Sehingga data akan dikelompokkan menjadi 3 *cluster*. Setelah mendapatkan nilai c , Langkah berikutnya yaitu menerapkan algoritma K-Means pada data nilai rata-rata mata pelajaran UN paket C PKBM Sanggar Puri dengan $c = 3$. Berikut ini merupakan hasil *cluster* nilai rata-rata mata pelajaran UN paket C PKBM Sanggar Puri menggunakan aplikasi *Rapid Miner*, seperti pada gambar 4 berikut.



Gambar 4. Proses *Clustering* dengan Aplikasi *Rapid Miner*

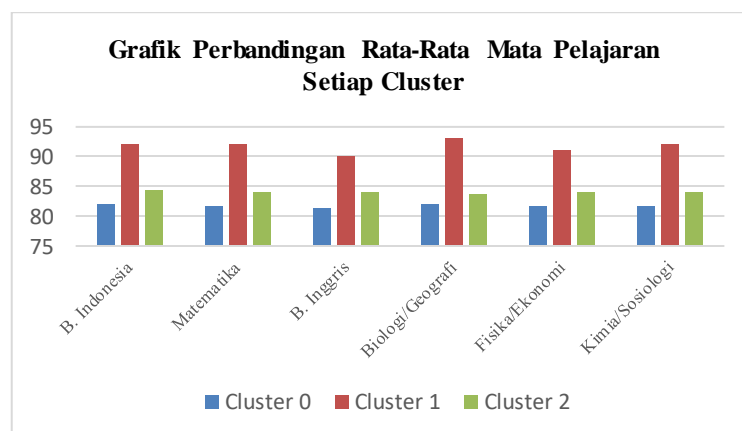
Di bawah ini gambar 5 merupakan Hasil *Cluster Model* dengan aplikasi *Rapid Miner*.

Cluster Model

```
Cluster 0: 57 items
Cluster 1: 1 items
Cluster 2: 42 items
Total number of items: 100
```

Gambar 5. Hasil *Cluster Model* dengan aplikasi *Rapid Miner*

Berdasarkan gambar 5 dapat dilihat jumlah anggota dari *cluster* 0 adalah 57 peserta didik, kemudian *cluster* 1 berjumlah 1 peserta didik, dan *cluster* 2 memiliki jumlah anggota *cluster* sebanyak 42 peserta didik. *Cluster* 0 merupakan *cluster* yang memiliki jumlah anggota terbanyak, kemudian disusul *cluster* 2, dan *cluster* 1 dengan jumlah anggota paling sedikit. Selanjutnya untuk mengetahui *cluster* tersebut termasuk kategori apa, maka dilakukan visualisasi menggunakan diagram batang seperti pada gambar 6.



Gambar 6. Grafik Perbandingan Rata-Rata Mata Pelajaran Setiap *Cluster*

Gambar 6 menunjukkan *cluster* 0 didominasi oleh rata-rata dengan nilai cenderung rendah dan jumlah anggota *cluster* yaitu 57 peserta didik. Lalu *cluster* 2 didominasi oleh nilai rata-rata sedang dengan jumlah anggota *cluster* yaitu 42 peserta

didik. Sedangkan *cluster* 1 merupakan kelompok nilai dengan rata-rata tinggi yang hanya berjumlah 1 peserta didik, sesuai pada tabel 3.

Tabel 3. Hasil Pengelompokkan Nilai Peserta Didik

| No | Bahasa Indonesia | Matematika | Bahasa Inggris | Biologi / Biografi | Fisika / Ekonomi | Kimia / Sosiologi | Cluster |
|-----|------------------|------------|----------------|--------------------|------------------|-------------------|-----------|
| 1 | 81 | 80 | 80 | 81 | 82 | 81 | Cluster 0 |
| 2 | 81 | 81 | 81 | 82 | 82 | 81 | Cluster 0 |
| 3 | 82 | 81 | 82 | 82 | 82 | 82 | Cluster 0 |
| ... | | | | | | | |
| 24 | 85 | 83 | 84 | 83 | 84 | 84 | Cluster 2 |
| 25 | 86 | 85 | 84 | 83 | 85 | 85 | Cluster 2 |
| 26 | 86 | 85 | 85 | 85 | 87 | 86 | Cluster 2 |
| ... | | | | | | | |
| 58 | 92 | 92 | 90 | 93 | 91 | 92 | Cluster 1 |
| 59 | 82 | 80 | 80 | 81 | 82 | 81 | Cluster 0 |
| 60 | 83 | 80 | 80 | 83 | 80 | 81 | Cluster 0 |
| ... | | | | | | | |
| 86 | 81 | 82 | 81 | 81 | 81 | 81 | Cluster 0 |
| 87 | 83 | 83 | 82 | 81 | 81 | 82 | Cluster 0 |
| 88 | 82 | 83 | 82 | 80 | 81 | 82 | Cluster 0 |
| ... | | | | | | | |
| 98 | 87 | 85 | 87 | 81 | 82 | 84 | Cluster 2 |
| 99 | 87 | 86 | 88 | 79 | 84 | 85 | Cluster 2 |
| 100 | 87 | 86 | 86 | 83 | 81 | 85 | Cluster 2 |

Dari hasil pengelompokkan yang telah dilakukan dapat diketahui bahwa terdapat kesenjangan jumlah anggota *cluster* 1 dengan *cluster* 0 dan 2. Hal tersebut menunjukkan bahwa proses KBM di PKBM Sanggar Puri belum merata dimana jumlah anggota *cluster* 1 atau kelompok dengan nilai rata-rata tinggi hanya berjumlah 1 orang. Artinya PKBM Sanggar Puri perlu melakukan strategi baru dalam proses belajar mengajar di PKBM Sanggar Puri agar hasil belajar peserta didik meningkat dan merata khususnya pada mata pelajaran UN. Tentunya hal ini menjadi perhatian khusus untuk mengatasi kesenjangan antara kelompok rata-rata nilai tinggi dengan kelompok rata-rata nilai sedang dan rendah.

3.5 Evaluation

Tahap evaluasi adalah mengukur performa hasil *cluster* nilai belajar peserta didik dengan jumlah *cluster* (c) = 3 menggunakan *Davies Bouldin Index* (DBI). Nilai DBI yang dihasilkan pada model ini menggunakan algoritma *K-Means* adalah 0.114 atau 88.4%, sesuai dengan gambar 7.

Davies Bouldin

Davies Bouldin: 0.114

Gambar 7. Nilai DBI dengan jumlah *cluster* (c) = 3

Kemudian untuk mengetahui apakah $c=3$ merupakan jumlah *cluster* paling optimal atau tidak maka dilakukan percobaan dengan menghitung nilai DBI dimulai dengan $c=2$ sampai $c=7$ seperti pada tabel 4.

Tabel 4. Perbandingan Jumlah *Cluster* (c) dengan DBI

| No | Jumlah <i>Cluster</i> (c) | <i>Davies Bouldin Index</i> (DBI) |
|----|---------------------------|-----------------------------------|
| 1 | 2 | 0,157 |
| 2 | 3 | 0,114 |
| 3 | 4 | 0,154 |
| 4 | 5 | 0,177 |
| 5 | 6 | 0,173 |
| 6 | 7 | 0,175 |

Berdasarkan tabel 4, nilai DBI paling kecil adalah saat jumlah *cluster* adalah 3. Semakin kecil nilai DBI atau mendekati 0 maka kemiripan data dalam suatu kelompok semakin berdekatan. Sebaliknya apabila semakin besar nilai DBI atau mendekati 1 maka kemiripan data dalam suatu kelompok semakin berjauhan. Sehingga terbukti $c=3$ merupakan jumlah *cluster* paling optimal.

4. KESIMPULAN

Algoritma *K-Means Clustering* dan metode *Elbow* dapat menghasilkan pengelompokan data yang baik. Algoritma *K-Means* mengelompokkan data berdasarkan kemiripan data yang berdasarkan jarak data dengan *centroid*. Hasil yang didapatkan dari proses clustering ini menghasilkan 3 *cluster* yaitu *cluster 0* merupakan kelompok rata-rata dengan nilai cenderung rendah dan jumlah anggota *cluster* yaitu 57 peserta didik. Lalu *cluster 2* merupakan kelompok nilai rata-rata sedang dengan jumlah anggota *cluster* yaitu 42 peserta didik. Sedangkan *cluster 1* merupakan kelompok nilai dengan rata-rata tinggi yang hanya berjumlah 1 peserta didik. Metode *Elbow* juga terbukti dapat digunakan dalam menentukan jumlah *cluster* paling ideal berdasarkan perbandingan SSE dengan jumlah *cluster* (c) yang memiliki sudut yang paling membentuk siku (*elbow*). Selain itu dari hasil perbandingan nilai *Davies Bouldin Index* (DBI) juga membuktikan bahwa $c=3$ merupakan jumlah *cluster* yang paling ideal karena nilai DBI yang paling mendekati 0 dibandingkan dengan jumlah *cluster* lain. Hasil akurasi menggunakan DBI pada penelitian ini adalah 88,4%. Dari hasil pemodelan yang telah didapat menggunakan *K-Means Clustering* maka hasil dari jumlah masing-masing *cluster* yang tidak seimbang perlu difokuskan untuk pemerataan proses belajar mengajar yang lebih baik. Antara lain dengan mengadakan kelas tambahan untuk persiapan Ujian Nasional (UN) dan merubah model pembelajaran yang lebih interaktif dan aktif. Kemudian saran untuk penelitian selanjutnya adalah menambah lebih banyak atribut dan menggunakan algoritma dan teknik evaluasi *cluster* yang lain sehingga informasi yang dihasilkan dapat lebih rinci dan optimal.

UCAPAN TERIMA KASIH

Peneliti ingin mengucapkan terima kasih kepada PKBM Sanggar Puri yang telah memberikan izin dalam pengambilan data, dan juga dosen terkait yang telah membimbing serta teman-teman yang telah memberikan dukungan dan semangat kepada peneliti.

DAFTAR PUSTAKA

- [1] A. Voutama, U. Enri, I. Maulana, dan E. Novalia, "Sosialisasi Literasi Digital Bagi Remaja dan Calistung Untuk Anak-Anak di Desa Telukbuyung Karawang," *J. Pemberdaya. Komunitas MH Thamrin*, vol. 4, no. 1, hal. 34–41, 2022, doi: 10.37012/jpkmht.v4i1.870.
- [2] R. R. Anderha, S. Maskar, dan U. T. Indonesia, "PENGARUH KEMAMPUAN NUMERASI DALAM MENYELESAIKAN MASALAH MATEMATIKA TERHADAP PRESTASI BELAJAR MAHASISWA PENDIDIKAN MATEMATIKA," *J. Ilm. Mat. Realis.*, vol. 2, no. 1, hal. 1–10, 2021.
- [3] A. Winarta dan W. J. Kurniawan, "Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, hal. 113–119, 2021.
- [4] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, hal. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.
- [5] E. A. Novia, W. I. Rahayu, dan C. Prianto, *Sistem Perbandingan Algoritma K-Means dan Naive Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan*. Kreatif, 2020.
- [6] R. Novianto dan L. Goeirmanto, "Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 1, hal. 85–95, 2019, doi: 10.35957/jatisi.v6i1.150.

-
- [7] V. Ramadhan dan A. Voutama, "Clustering Menggunakan Algoritma K-Means Pada Penyakit ISPA di Puskesmas Kabupaten Karawang," vol. 4, hal. 462–473, 2022.
- [8] T. Purwanti, W. Ramdhan, dan Santoso, "PENERAPAN METODE KLASTERISASI K-MEANS UNTUK STRATEGI PROMOSI PADA SMK TAMANSISWA SUKADAMAI," *JUTSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 1, no. 2, hal. 141–146, 2021.
- [9] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, hal. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [10] R. Ananda dan A. Z. Yamani, "Penentuan Centroid Awal K-Means pada Proses Clustering Data Evaluasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, hal. 544–550, 2021.
- [11] M. Rafi Muttaqin, T. Iman Hermanto, M. Agus Sunandar, P. Studi Teknik Informatika, dan S. Tinggi Teknologi Wastukencana, "Penerapan K-Means Clustering Dan Cross-Industry Standard Process for Data Mining (Crisp-Dm) Untuk Mengelompokan Penjualan Kue," vol. 19, no. 1, hal. 38–53, 2022.
- [12] A. Yoga Pratama, Y. Umaidah, dan A. Voutama, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, hal. 897–910, 2021.
- [13] Y. Christian dan K. O. Y. R. Qi, "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Early Stage Startup dengan Menggunakan CRISP-DM," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, hal. 966–973, 2022, doi: 10.30865/jurikom.v9i4.4486.
- [14] D. L. Rianti, Y. Umaidah, dan A. Voutama, "Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 6, no. 1, hal. 98, 2021, doi: 10.30998/string.v6i1.9993.
- [15] D. Jollyta, M. Siddik, H. Mawengkang, dan S. Efendi, *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer*. Deepublish, 2021.
- [16] D. Jollyta, S. Efendi, M. Zarlis, dan H. Mawengkang, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, hal. 918, 2019, doi: 10.30645/senaris.v1i0.100.
- [17] N. H. Harani, C. Prianto, dan F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python," *J. Manaj. Inform.*, vol. 10, no. 2, hal. 133–146, 2020, doi: 10.34010/jamika.v10i2.2683.
- [18] W. M. Baihaqi, K. Indartono, dan S. Banat, "Penerapan Teknik Clustering Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 2, hal. 243–248, 2019, doi: 10.31294/p.v21i2.6149.