



## Klasterisasi Tingkat Pemahaman Siswa Pada Mata Pelajaran Kimia Dengan Algoritma K-Means

Muhamad Rizal Gufran B<sup>1</sup>, Arif Dwi Nugroho<sup>2</sup>, Irham Asdurroh<sup>3</sup>, Ariyo Putra P<sup>4</sup>, Tedy Setiadi<sup>5</sup>

<sup>1,2,3,4,5</sup> Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Email : <sup>1</sup> muhamad2000018329@webmail.uad.ac.id

### INFO ARTIKEL

#### Sejarah Artikel:

Diterima: 27 Januari 2023  
Direvisi: 13 Februari 2023  
Disetujui: 3 Maret 2023  
Dipublikasi: 13 Maret 2023

#### Kata Kunci:

Data Mining  
Klustering  
K-Means

### ABSTRAK

Kimia merupakan mata pelajaran yang sebagian siswa merasa kesukasan dalam menerima materi, sehingga bagi guru mata pelajaran kimia hal ini merupakan sebuah masalah dalam memberikan materi karena tingkat pemahaman siswa yang tidak merata, hal ini tentu saja akan menyebabkan tujuan pembelajaran yang tidak dapat dicapai sesuai rencana studi semester. Tujuan penelitian ini dilakukan adalah untuk mengelompokkan tingkat pemahaman siswa pada mata pelajaran kimia berdasarkan nilai PAS/PAT dan Rapor. Metode yang akan digunakan yaitu Clustering dengan algoritma K-Means, untuk menentukan jumlah cluster yang optimal dengan metode elbow. Setelah dilakukan penelitian ditemukan jumlah cluster yang optimal sebanyak 3 cluster, dimana cluster 0 merupakan kelompok siswa yang memiliki tingkat pemahaman sangat Baik, Cluster 1 merupakan kelompok tingkat pemahaman baik dan Cluster 2 merupakan kelompok tingkat pemahaman kurang. Dari 133 siswa yang digunakan sebagai dataset, masing-masing didapatkan hasil untuk cluster 0 sebanyak 78 siswa, cluster 1 sebanyak 54 siswa dan cluster 2 sebanyak 1 siswa. Siswa yang masuk pada cluster 2 perlu mendapatkan perhatian khusus dari guru terkait sehingga dapat memahami pelajaran dengan baik.

### Keyword:

Data Mining  
Clustering  
K-Means

### DOI Artikel:

10.35891/explorit.v14i2.3665

### ABSTRACT

Chemistry is a subject in which some students find it difficult to accept the material. This poses a problem for chemistry teachers in providing materials, as it leads to inconsistent student comprehension and learning goals that cannot be achieved by the curriculum. The purpose of this research was to classify students' level of understanding in chemistry based on PAS/PAT scores and report cards. The method to be used is clustering with the K-Means algorithm, to determine the optimal number of clusters with the elbow method. After conducting the research, it was found that the optimal number of clusters was 3 clusters, where cluster 0 was a group of students who had a very good level of understanding, Cluster 1 was a group with a good level of understanding and Cluster 2 was a group with a low level of understanding. Of the 133 students used as a dataset, each obtained results for cluster 0 of 78 student, cluster 1 of 54 students and cluster 2 of 1 student. Students who enter cluster 2 need special attention from the relevant teacher so that they can understand the lesson well.

@2022 diterbitkan oleh Prodi Teknik Informatika Universitas Yudharta Pasuruan

## 1. Pendahuluan

MAN 1 Baubau merupakan Madrasah Aliyah Negeri yang berada di Kota Baubau. Mata pelajaran yang diberikan oleh MAN 1 Baubau mencakup seluruh mata pelajaran wajib yang sesuai dengan kurikulum yang telah berlaku serta pelajaran tambahan dan nilai-nilai dalam agama Islam. Kimia merupakan salah satu mata pelajaran wajib yang ada di MAN 1 Baubau.

Kimia merupakan suatu mata pelajaran yang mempelajari tentang materi, perubahan materi, dan energi yang menyertainya. Pada mata pelajaran kimia memfokuskan pada penguasaan konsep. Dalam proses belajar, konsep adalah hal-hal yang harus dipelajari, dikuasai dan dipahami oleh seorang siswa. Kimia adalah mata pelajaran yang cukup kompleks dengan banyak konsep abstrak untuk dipahami dan dipelajari oleh siswa. Pada mata pelajaran kimia setiap materinya saling berkaitan antara materi yang satu dengan materi lainnya. Apabila siswa tidak bisa memahami materi sebelumnya maka akan kesulitan untuk memahami materi berikutnya. Sehingga menjadikan siswa cenderung kesulitan untuk memahami pelajaran kimia. Bagi guru pengampu mata pelajaran kimia hal ini merupakan suatu masalah dalam memberikan sebuah materi karena pemahaman siswa yang tidak merata, hal ini tentu saja akan menyebabkan tujuan pembelajaran yang tidak dapat dicapai sesuai dengan rencana studi semester.

Teknologi data mining, dapat digunakan untuk menganalisis tingkat pemahaman siswa. Analisis ini dilakukan untuk mengelompokkan siswa sesuai dengan kemampuan dalam memahami dan menguasai materi pelajaran kimia. Pengelompokkan tingkat pemahaman siswa menggunakan teknik atau metode clustering dilakukan dengan cara membagi kelompok siswa menjadi sub kelompok yang disebut cluster. Clustering adalah metode pada data mining yang digunakan untuk menganalisis data supaya lebih akurat saat menyelesaikan masalah pengelompokan data atau membagi sekumpulan data menjadi himpunan bagian. Tujuan clustering adalah mengelompokkan data ke dalam suatu kelompok kemudian hubungan antar anggota cluster yang sama menjadi lebih kuat, sedangkan hubungan antar anggota cluster yang berbeda menjadi semakin lemah. Objek pada suatu cluster mempunyai

kemiripan karakteristik tetapi mempunyai karakteristik yang berbeda dengan objek pada cluster lainnya. Sehingga Clustering digunakan untuk menentukan kelompok atau cluster yang tidak diketahui pada suatu data.

Penelitian ini menggunakan teknik atau metode k-means clustering untuk melakukan pengelompokan tingkat pemahaman siswa pada pelajaran kimia. Dengan menggunakan teknik clustering k-means diharapkan guru maupun pihak sekolah bisa mendata siswa dengan tingkat pemahamannya masing-masing serta dapat melakukan proses pengajaran dengan cara yang tepat sehingga siswa bisa memahami pelajaran kimia dengan baik. Fokus penelitian ini pada pengelompokan tingkat pemahaman siswa di MAN 1 BAUBAU pada pelajaran kimia dengan menggunakan metode atau teknik clustering k-means, dimana siswa dapat dikelompokkan menjadi tiga cluster yaitu tingkat pemahaman yang sangat baik, baik, dan kurang, berdasarkan nilai PAS/PAT dan rapor.

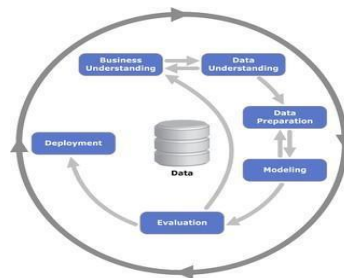
## 2. Kajian Teori

### 2.1 Data Mining

Data mining merupakan proses penggalian informasi yang berguna pada database yang besar dan harus diekstraksi sehingga menjadi informasi yang baru dan membantu untuk melakukan pengambilan keputusan. Data mining merupakan suatu proses menganalisa data dari bermacam-macam sumber dan menggabungkannya menjadi suatu pengetahuan atau informasi yang penting untuk menaikkan keuntungan, mengurangi biaya, atau bahkan keduanya. [7]

### 2.2 CRISP-DM

CRISP-DM digunakan sebagai sebuah pemecah masalah dalam bisnis maupun penelitian. Tahapan Metode CRISP-DM yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment[8]. Tahapan CRISP-DM dapat dilihat pada gambar 1.



Gambar 1. Tahapan CRISP-DM

#### 2.2.1 Business Understanding

Pada tahap ini dilakukan untuk memahami kebutuhan dan tujuan dari perspektif bisnis, kemudian menginterpretasikan pengetahuan dalam bentuk mendefinisikan masalah data mining serta mendefinisikan rencana dan strategi untuk memperoleh tujuan data mining.[9]

#### 2.2.2 Data Understanding

Tahap ini merupakan proses mengumpulkan data apa saja yang dimiliki dan dibutuhkan. Proyek analisis data dimulai dengan penemuan data yang sudah ada, kemudian mengarahkan analyst untuk menggali knowledge yang terdapat dalam kumpulan data tersebut. Tipe dari data akan menentukan tujuan dan jenis algoritma pada data mining yang ingin dicapai.[1]

#### 2.2.3 Data Preparation

Pada tahapan ini membuat dataset akhir dari data mentah. Beberapa hal yang akan dikerjakan antara lain Data Cleaning (Pembersihan Data), Data Selection (Pemilihan Data) pada record dan atribut, serta Data Transformation (Transformasi Data) sebagai input pada tahap pemodelan. [9]

#### 2.2.4 Modeling

Modeling dilakukan untuk menciptakan model prediktif maupun deskriptif. Pada tahapan ini ditentukan teknik data mining, tools data mining beserta algoritma data mining yang akan digunakan dengan menggunakan metode statistika dan machine learning. Kemudian, langkah selanjutnya adalah menerapkan teknik serta algoritma data mining tersebut kepada data dengan menggunakan tools. Pemodelan yang sering dilakukan antara lain *klasifikasi, evaluasi, ranking, clustering, finding relation, dan karakterisasi*. [8]

#### 2.2.5 Evaluation

Menafsirkan hasil data mining yang didapatkan pada tahapan pemodelan. Evaluasi ini dilakukan terhadap model yang diimplementasikan pada tahap pemodelan, sehingga model yang telah ditentukan dapat sesuai dengan tujuan yang ingin dicapai pada tahap business understanding.[8]

#### 2.2.6 Deployment

Pada tahap ini akan dilakukan dengan membuat sebuah laporan, jurnal atau artikel menggunakan model yang diperoleh.[1]

### 2.3 Clustering

Clustering merupakan suatu teknik mengelompokkan data. Clustering adalah suatu proses pengelompokkan data menjadi beberapa kelompok sehingga data dalam satu kelompok memiliki tingkat kemiripan maksimum dan data antar kelompok memiliki kemiripan minimum. Clustering merupakan proses mengelompokkan sekumpulan objek data atau sekumpulan cluster sehingga objek dalam satu kelompok sangat mirip, tetapi berbeda dengan objek di kelompok lain.[4]

### 2.4 K-Means

K-means merupakan sebuah algoritma yang memberikan nilai cluster secara acak, kemudian nilai tersebut akan menjadi titik pusat cluster yang disebut centroid. K-means adalah sebuah algoritma clustering dengan ide sederhana yaitu meminimalkan SSE (Sum of Squared Error) antara objek data dengan sejumlah centroid.[10]

### 2.5 Elbow

Elbow adalah suatu metode yang diterapkan pada algoritma K-Means untuk menentukan jumlah cluster yang optimal.[2]

## 3. Metodologi Penelitian

Tahap-tahap metodologi penelitian yang akan dilakukan sebagai berikut

3.1 Business Understanding

Tahapan ini menentukan dan memahami tujuan dari penelitian yang akan dilakukan.

3.2 Data Understanding

Tahap ini memahami dan menentukan dataset yang digunakan dalam penelitian.

3.3 Data Preparation

Tahap ini dilakukan untuk mempersiapkan data mentah yang akan diproses

3.4 Modeling

Pada tahap ini berisi hasil pemodelan dari hasil penelitian.

3.5 Evaluation

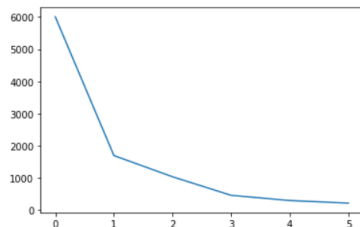
Tahapan ini menjelaskan hasil pengujian pada penelitian

3.6 Deployment

Tahap ini akan dilakukan dengan membuat laporan atau artikel jurnal menggunakan model yang diperoleh

4. Hasil Uji Coba Dan Pembahasan

Setelah melakukan pengujian menggunakan metode elbow didapatkan jumlah cluster yang optimal adalah 3. Dapat dilihat dari titik pada grafik yang penurunan inersianya tidak lagi signifikan seperti pada gambar 2, maka jumlah cluster yang optimal adalah 3.



Gambar 2. Hasil Visualisasi Metode Elbow

4.1 Business Understanding (Pemahaman Bisnis)

Penelitian ini dilakukan untuk mengelompokkan tingkat pemahaman siswa terhadap mata pelajaran kimia berdasarkan nilai PAS/PAT dan Rapor.

4.2 Data Understanding (Pemahaman Data)

Penelitian ini memakai data yang berasal dari dataset nilai raport siswa kelas 12 MAN 1 Baubau pada pelajaran Kimia. Jumlah Data yang digunakan sebanyak 133 data sesuai dengan jumlah siswa. Dataset ini memiliki 9 atribut yaitu No, NISN, Nama, L/P, Harian PAS/PAT Rapor, Predikat dan Deskripsi. Adapun datasetnya dapat dilihat pada tabel 1.

Tabel 1. Dataset Nilai Raport Siswa MAN 1 Baubau

No	NISN	Nama	L/ P	Harian	PAS/ PAT	Rapor	Predikat
1	0068757450	AFIFAH ULYA QURRATA AINI	P	80	80	80	B
2	0064551214	AGHISTA ZARA SYAVITRI	P	80	80	80	B
3	0071747289	ANISA SARWAN	P	78	78	78	C
4	0078066468	AYU FADILAH	P	78	78	78	C
5	0075951831	CEIS AGISTINA	P	78	78	78	C
6	0078010944	CHINTANY AYSKHA FAYRANA SAFRUDDIN	P	80	80	80	B
7	0067196469	DIAN AL FAIZAH	P	75	75	75	C
8	0071248270	FAJAR BAKTI SUPRAPTO	L	80	80	80	B
9	0074302038	L.M RIZKI FEBRIAWAN RAHMAD	L	78	78	78	C
...	...	...	...	...	...	...	...
130	3068557424	WA MIRA	P	88	87	88	B
131	0049494903	WA NILA	P	88	87	88	B

132	0038073220	WA ODE RISTA DEWI	P	88	87	88	B
133	0034839626	ZAMIRUN	L	80	80	80	C

### 4.3 Data Preparation (Persiapan Data)

#### 4.3.1 Impor Data

Pada proses ini dilakukan import dataset nilai raport siswa dengan format CSV yang terdiri dari 133 record dan 9 atribut.

#### 4.3.2 Transformation

Pada proses ini digunakan untuk mengubah nama atribut pada dataset sehingga dapat diolah. Hasil setelah transformasi dapat dilihat pada tabel 2.

Tabel 2. Hasil Transformation Dataset

No	NISN	Nama	Jenis_Kelamin	Nilai_Harian	Nilai_PAS	Nilai_Rapor	Predikat
1	0068757450	AFIFAH ULYA QURRATA AINI	P	80	80	80	B
2	0064551214	AGHISTA ZARA SYAVITRI	P	80	80	80	B
3	0071747289	ANISA SARWAN	P	78	78	78	C
4	0078066468	AYU FADILAH	P	78	78	78	C
5	0075951831	CEIS AGISTINA	P	78	78	78	C
6	0078010944	CHINTANY AYSKHA FAYRANA SAFRUDDIN	P	80	80	80	B
7	0067196469	DIAN AL FAIZAH	P	75	75	75	C
8	0071248270	FAJAR BAKTI SUPRAPTO	L	80	80	80	B
9	0074302038	L.M RIZKI FEBRIAWAN RAHMAD	L	78	78	78	C
...	...	...	...	...	...	...	...
130	3068557424	WA MIRA	P	88	87	88	B
131	0049494903	WA NILA	P	88	87	88	B
132	0038073220	WA ODE RISTA DEWI	P	88	87	88	B
133	0034839626	ZAMIRUN	L	80	80	80	C

#### 4.3.3 Selection

Proses ini untuk menentukan apa saja atribut yang akan digunakan dalam proses clustering. Hasil atribut setelah melakukan selection ditunjukkan pada tabel 3.

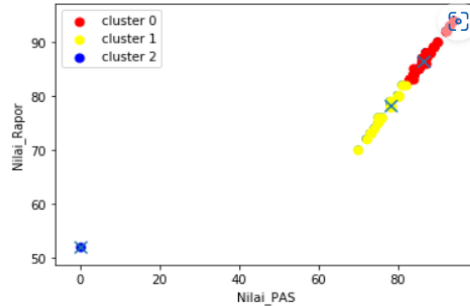
Tabel 3. Hasil Selection Dataset

NISN	Nama	Nilai_PAS	Nilai_Rapor
0068757450	AFIFAH ULYA QURRATA AINI	80	80
0064551214	AGHISTA ZARA SYAVITRI	80	80
0071747289	ANISA SARWAN	78	78
0078066468	AYU FADILAH	78	78
0075951831	CEIS AGISTINA	78	78
0078010944	CHINTANY AYSKHA FAYRANA SAFRUDDIN	80	80
0067196469	DIAN AL FAIZAH	75	75
0071248270	FAJAR BAKTI SUPRAPTO	80	80
0074302038	L.M RIZKI FEBRIAWAN RAHMAD	78	78
...	...	...	...
3068557424	WA MIRA	87	88

0049494903	WA NILA	87	88
0038073220	WA ODE RISTA DEWI	87	88
0034839626	ZAMIRUN	80	80

4.4 Modelling (Pemodelan)

Pemodelan clustering dengan algoritma K-Means menggunakan Jupyter notebook, hasil pemodelan clustering dapat dilihat pada gambar 3.



Gambar 3. Visualisasi Hasil Clustering

4.5 Evaluation (Pengujian)

Setelah melakukan pemodelan dapat menentukan titik Centroid dari setiap cluster dengan melihat letak titik pusat cluster yang diberi tanda silang berwarna hijau atau dapat diketahui dengan memberikan perintah “print(cluster)” lalu diperoleh hasil seperti pada gambar 4.

```
#melihat pusat cluster 1-3
print(cluster)

[[ 86.42307692  86.47435897]
 [ 78.2037037  78.25925926]
 [ 0.          52.          ]]
```

Gambar 4. Centroid Cluster

Jumlah anggota setiap cluster dapat dilihat dengan menggunakan perintah “tabel['cluster'].value\_counts()”. Hasil dari perintah tersebut dapat dilihat pada gambar 5.

```
: tabel['cluster'].value_counts()

: 0    78
  1    54
  2     1
Name: cluster, dtype: int64
```

Gambar 5. Jumlah Anggota Setiap Cluster

Berdasarkan dua perintah sebelumnya, maka hasilnya dapat dimodelkan seperti pada tabel 4.

Tabel 4. Centroid dan Jumlah Siswa

Cluster	Centroid		Jumlah Siswa
	Nilai PAS	Nilai Rapor	
0	86.42307692	86.47435897	78
1	78.2037037	78.25925926	54
2	0	52	1

Melihat skor Centroid tiap cluster pada table di atas, kita dapat menganalisis atau mengklasifikasikan cluster 0 sebagai cluster pemahaman siswa sangat baik sebanyak 78 siswa , cluster 1 baik sebanyak 54 siswa dan cluster 2 kurang baik sebanyak 1 orang siswa, sehingga dapat dikatakan kelompok 2 adalah kelompok siswa perlu perhatian khusus agar dapat memahami materi dengan baik, agar proses pembelajaran berjalan dengan lancar serta tujuan pembelajaran dapat tercapai. Hasil pengelompokan tiap siswa dapat dilihat pada tabel 5.

Tabel 5. Hasil Clustering Dataset

NISN	Nama	Nilai_PAS	Nilai_Rapor	Cluster
0068757450	AFIFAH ULYA QURRATA AINI	80	80	1
0064551214	AGHISTA ZARA SYAVITRI	80	80	1
0071747289	ANISA SARWAN	78	78	1
0078066468	AYU FADILAH	78	78	1
0075951831	CEIS AGISTINA	78	78	1
0078010944	CHINTANY AYSKHA FAYRANA SAFRUDDIN	80	80	1
0067196469	DIAN AL FAIZAH	75	75	1
0071248270	FAJAR BAKTI SUPRAPTO	80	80	1
0074302038	L.M RIZKI FEBRIAWAN RAHMAD	78	78	1
...	...	...	...	
3068557424	WA MIRA	87	88	0
0049494903	WA NILA	87	88	0
0038073220	WA ODE RISTA DEWI	87	88	0
0034839626	ZAMIRUN	80	80	1

## 5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka jumlah kelompok yang digunakan untuk mengelompokkan tingkat pemahaman siswa kelas 12 MAN 1 Baubau pada mata pelajaran kimia adalah 3 kelompok. Dimana Kelompok 0 merupakan kelompok siswa dengan tingkat pemahaman Baik, Kelompok 1 merupakan kelompok Baik dan Kelompok 2 berada pada tingkat Kurang Baik. Dari 133 siswa yang dijadikan dataset, masing-masing kelompok diperoleh kelompok 0 sebanyak 78 siswa, kelompok 1 sebanyak 54 siswa, dan kelompok 2 sebanyak 1 siswa. Berdasarkan hasil pembagian kelompok, anak-anak kelompok 2 memerlukan perhatian khusus dari guru agar mereka dapat memahami pelajaran dengan baik.

## 6. Daftar Pustaka

- [1] Msy Aulia Hasanah , Sopian Soim 2, Ade Silvia Handayani 3. 2021. "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir",
- [2] Rezqiwati Ishak, Amiruddin. 2022. "Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode K-Means".
- [3] Austin Almayda , Sudin Saepudin. 2021. "Penerapan Data Mining K-Means Clustering Untuk Mengelompokkan Berbagai Jenis Merk Smartphone"
- [4] Magdalena Simanjuntak, Ediman Manik, Tri Supratman. 2018. PENERAPAN DATA MINING PENGELOMPOKAN KEJAHATAN ELEKTRONIK SESUAI UU ITE DENGAN MENGGUNAKAN METODE CLUSTERING
- [5] Tedy Setiadi, Jamaludin. 2018. "Penerapan Klasifikasi Bayes Untuk Memprediksi Jenis Latihan Siswa Pencak Silat (Studi Kasus Pencak Silat PSHT)"
- [6] Mujib Ridwan, Hadi Suyono, dan M. Sarosa. (2013). " Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier".
- [7] Adjie Kukuh Wahyudi, Noor Azizah , Heru Saputro. 2022. DATA MINING KLASIFIKASI KEPRIBADIAN SISWA SMP NEGERI 5 JEPARA MENGGUNAKAN METODE DECISION TREE ALGORITMA C4.5
- [8] Annisa Paramitha Fadillah. 2015. Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)
- [9] Msy Aulia Hasanah, Sopian Soim , Ade Silvia Handayani. 2021. Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir
- [10] Diah Ayu Maulida Wati, Diah Puspitasari, Esty Purwaningsih. 2019. Metode Clustering Pada Model Algoritma K-Means Untuk Pemilihan Alat Kontrasepsi.