

BAB IV

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan di SMA Negeri 1 Pagelaran, satu-satunya SMA Negeri di Kecamatan Pagelaran, Kabupaten Pringsewu. Data yang digunakan dalam penelitian ini mencakup *dataset* akademik dari tahun 2020 hingga 2023. *Dataset* tersebut terdiri dari 982 data siswa, dengan atribut meliputi jenis kelamin, jurusan, nilai akademik, nilai non akademik, dan status alumni. Pengumpulan data dilakukan pada tanggal 9 Desember 2024. Tujuan dari penelitian ini adalah untuk menganalisis pola data akademik menggunakan algoritma *K-Means Clustering*. Melalui analisis ini, diharapkan dapat disusun strategi yang efektif untuk meningkatkan mutu sekolah, baik dari segi akademik maupun non-akademik. Dalam penelitian ini menggunakan *Software Rapidminer*.

4.1 Data Selection

Dataset yang digunakan dalam penelitian ini adalah data akademik siswa di SMA Negeri 1 Pagelaran yang mencakup periode tahun 2020 hingga 2023. *Dataset* ini terdiri dari 981 baris data mengandung berbagai informasi, seperti nama, jenis kelamin, jurusan, nilai akademik, nilai non akademik, serta status alumni. Langkah pertama dalam pengolahan data adalah pembersihan *dataset* dengan membuang data yang tidak konsisten, seperti data duplikat, data yang tidak lengkap, atau data dengan anomali. Proses ini bertujuan untuk memastikan kualitas dan keakuratan *dataset* sebelum dilakukan analisis lebih lanjut. Setelah pembersihan, data mentah diubah atau ditransformasikan dari format awalnya menjadi format yang lebih sesuai untuk digunakan dalam proses pengelompokan (*clustering*). Proses transformasi ini melibatkan normalisasi nilai, seperti mengubah rentang nilai akademik dan non-akademik agar berada dalam skala yang seragam, sehingga algoritma dapat memprosesnya dengan lebih efektif. Selain itu, atribut kategori, seperti jenis kelamin dan jurusan, dikodekan menggunakan metode *one-hot encoding* untuk memastikan data dapat dimengerti oleh algoritma *machine*

learning. Dari hasil transformasi tersebut, lima atribut utama diseleksi untuk digunakan dalam penelitian ini, yaitu jenis kelamin, jurusan, nilai akademik, nilai non-akademik, dan status alumni. Kelima atribut ini dipilih karena memiliki relevansi signifikan terhadap pengelompokan siswa berdasarkan karakteristik akademik dan non-akademik mereka. Data yang telah selesai diproses kemudian disimpan dalam suatu berkas khusus yang terpisah dari basis data operasional sekolah. Penyimpanan ini dilakukan untuk menjaga integritas data asli sekaligus mempermudah proses analisis *data mining* menggunakan algoritma *K-Means Clustering*. Dengan langkah-langkah transformasi dan seleksi atribut yang telah dilakukan, *dataset* yang dihasilkan siap untuk dianalisis guna mengidentifikasi pola-pola penting yang dapat mendukung penyusunan strategi peningkatan mutu pendidikan di SMA Negeri 1 Pagelaran. Statistik data yang digunakan ada pada gambar 4.1.

| Name | Type | Missing | Statistics | | |
|--------------------|---------|---------|------------------------------|--------------------------------|--|
| NAMA | Nominal | 0 | Least Zikri Ramadhani (1) | Most DHIO RAFFA AGIASTA ... | Values DHIO RAFFA AGIASTA (3), RADITHYA MUHAMMAD FAQIH (3), ...[698 more] |
| Jenis Kelamin | Nominal | 0 | Least laki (356) | Most perempuan (625) | Values perempuan (625), laki (356) |
| JURUSAN | Nominal | 0 | Least IPS (81) | Most IPA (593) | Values IPA (593), IPS (307), ...[1 more] |
| Nilai Akademik | Real | 0 | Min 72 | Max 88 | Average 79.222 |
| Nilai Non Akademik | Real | 1 | Min 72.060 | Max 88 | Average 79.151 |
| STATUS ALUMNI | Nominal | 0 | Least Kuliah (475) | Most Bekerja (506) | Values Bekerja (506), Kuliah (475) |

Gambar 4.1 Statistik seleksi data yang digunakan

Tabel 4. 1 Sembilan Baris Pertama Data *Selection*

| Nama Siswa | Jenis Kelamin | Jurusan | Nilai Akademik | Nilai Non Akademik | Status Alumni |
|--------------------------------|---------------|---------|----------------|--------------------|---------------|
| Abel Tri Yunita | laki | ipa | 77.33 | 77.20 | Bekerja |
| Ade lia Dwi Ous0ita Ningrum | perempuan | ipa | 77.80 | 78.07 | Bekerja |
| Adi lia Dwi Wiratma | perempuan | ipa | 80.27 | 80.13 | Bekerja |
| Adinda Intania Iskandar | perempuan | ipa | 77.87 | 77.60 | Bekerja |
| Aditya Dwi Mau l ana | laki | ipa | 76.73 | 77.20 | Kuliah |
| Afifah Taskia | perempuan | ipa | 83.47 | 83.07 | Bekerja |

| | | | | | |
|----------------------|-----------|-----|-------|-------|---------|
| Ages SeOrina | laki | ipa | 81.47 | 81.07 | Kuliah |
| Ajeng Fibri Orastika | perempuan | ipa | 81.00 | 80.67 | Bekerja |
| Akbar Kurniawan | laki | ips | 78.33 | 77.73 | Kuliah |

4.2 Pre-processing / Cleaning

Dilakukannya proses *cleaning* data untuk menghilangkan data yang tidak konsisten, atau menghapus atribut yang tidak diperlukan untuk meminimalisir terjadinya *missing/error* pada saat proses *clustering* di *RapidMiner*. Terdapat 5 atribut yang akan digunakan yaitu Jenis Kelamin, Jurusan, Nilai Akademik, Nilai Non Akademik dan Status Alumni. Pada tahapan ini data akan dilakukan *cleaning* atau pembersihan data. Pada tahapan ini data akan dilakukan *cleaning* atau pembersihan data.

4.2.1 Cleaning Data (Missing Value dan Duplikat Data)

a. Statistik Data

Statistik data digunakan untuk menghapus spasi di awal maupun di akhir yang ada pada *dataset*, sehingga data dapat dikelompokkan dengan tepat. Pada gambar 4.2 dapat dilihat bahwa data yang belum di Statistik data memiliki 3 jurusan yaitu IPA, IPS dan IPS yang seharusnya adalah IPA dan IPS saja. Hasil fungsi trim dapat dilihat pada gambar 4.3.

| Name | Type | Missing | Statistics | Filter (6 / 6 attributes) | Search for Attributes |
|--------------------|------------|---------|---|---------------------------|-----------------------|
| NAMA | Polynomial | 0 | Least: Zikri Ramadhani (1) Most: DHIO RAFFA AGIASTA ... Values: DHIO RAFFA AGIASTA (3), RADITHYA MUHAMMAD FAQIH (3), ... [696 more] | | |
| Jenis Kelamin | Polynomial | 0 | Least: laki (356) Most: perempuan (625) Values: perempuan (625), laki (356) | | |
| JURUSAN | Polynomial | 0 |  Least: IPS (81) Most: IPA (593) Values: IPA (593), IPS (307), IPS (81) Details... | | |
| Nilai Akademik | Real | 0 | Min: 72 Max: 88 Average: 79.222 | | |
| Nilai Non Akademik | Real | 1 | Min: 72.060 Max: 88 Average: 79.151 | | |
| STATUS ALUMNI | Polynomial | 0 | Least: Kuliah (475) Most: Bekerja (506) Values: Bekerja (506), Kuliah (475) | | |

Gambar 4.2 Statistik *Dataset* sebelum Menghapus Spasi

| Name | Type | Missing | Statistics | | |
|----------------------|------------|---------|------------------------------|--------------------------------|---|
| ▼ NAMA | Polynomial | 0 | Least Zikri Ramadhani (1) | Most DHIO RAFFA AGIASTA ... | Values DHIO RAFFA AGIASTA (3), RADITHYA MUHAMMAD FAQIH (3), ... [698 more] |
| ▼ Jenis Kelamin | Polynomial | 0 | Least laki (356) | Most perempuan (625) | Values perempuan (625), laki (356) |
| ▼ JURUSAN | Polynomial | 0 | Least IPS (388) | Most IPA (593) | Values IPA (593), IPS (388) |
| ▼ STATUS ALUMNI | Polynomial | 0 | Least Kuliah (475) | Most Bekerja (506) | Values Bekerja (506), Kuliah (475) |
| ▼ Nilai Akademik | Real | 0 | Min 72 | Max 88 | Average 79.222 |
| ▼ Nilai Non Akademik | Real | 1 | Min 72.060 | Max 88 | Average 79.151 |

Gambar 4.3 Statistik *Dataset* setelah Menghapus Spasib. *Missing Value*

Missing value adalah data yang hilang atau kosong dalam sebuah *dataset*. Artinya, tidak ada nilai yang tercatat untuk suatu *variabel* pada entri tertentu.

| Name | Type | Missing | Statistics | | |
|----------------------|------------|---------|------------------------------|--------------------------------|---------------------------------------|
| ▼ NAMA | Polynomial | 0 | Least Zikri Ramadhani (1) | Most DHIO RAFFA AGIASTA ... | Values DHIO RAFFA AGIASTA (3), RAC |
| ▼ Jenis Kelamin | Polynomial | 0 | Least laki (356) | Most perempuan (625) | Values perempuan (625), laki (356) |
| ▼ JURUSAN | Polynomial | 0 | Least IPS (388) | Most IPA (593) | Values IPA (593), IPS (388) |
| ▼ STATUS ALUMNI | Polynomial | 0 | Least Kuliah (475) | Most Bekerja (506) | Values Bekerja (506), Kuliah (475) |
| ▼ Nilai Akademik | Real | 0 | Min 72 | Max 88 | Average 79.222 |
| ▼ Nilai Non Akademik | Real | 1 | Min 72.060 | Max 88 | Average 79.151 |

Gambar 4.4 Statistik *Dataset* Sebelum *Missing Value*

| Row... ↑ | NAMA | Jenis Kelamin | JURUSAN | STATUS... | Nilai Ak... | Nilai No... |
|----------|----------------------|---------------|---------|-----------|-------------|-------------|
| 1 | ROFIIF SYAMS SHOBAAH | laki | IPA | Bekerja | 72.670 | ? |

Gambar 4.5 Data yang terdapat *missing value*

Dataset yang digunakan dalam penelitian ini ditemukan satu *missing value* atau nilai kosong pada atribut Nilai Non Akademik, yang memiliki tipe data *Real*. *Missing value* tersebut menunjukkan adanya data yang tidak lengkap pada atribut tersebut, sehingga diperlukan langkah penanganan untuk memastikan *dataset* dapat

digunakan secara optimal dalam proses analisis. Langkah yang dilakukan untuk mengatasi masalah ini adalah proses "*Replace Missing Value*", yaitu mengganti nilai kosong pada atribut Nilai Non Akademik dengan nilai tertentu. Proses ini bertujuan untuk melengkapi data sehingga analisis dapat dilakukan tanpa adanya kendala akibat data yang hilang. Hasil dari proses "*Replace Missing Value*" dapat dilihat pada Gambar 4.4, yang menunjukkan bahwa nilai kosong pada atribut tersebut telah digantikan dengan nilai yang sesuai. Dengan demikian, dataset telah lengkap dan siap untuk digunakan dalam tahap analisis berikutnya.

| Name | Type | Missing | Statistics | | |
|----------------------|------------|---------|------------------------------|--------------------------------|--|
| ▼ NAMA | Polynomial | 0 | Least Zikri Ramadhani (1) | Most DHIO RAFFA AGIASTA ... | Values DHIO RAFFA AGIASTA (3), RADITI |
| ▼ Jenis Kelamin | Polynomial | 0 | Least laki (356) | Most perempuan (625) | Values perempuan (625), laki (356) |
| ▼ JURUSAN | Polynomial | 0 | Least IPS (388) | Most IPA (593) | Values IPA (593), IPS (388) |
| ▼ STATUS ALUMNI | Polynomial | 0 | Least Kuliah (475) | Most Bekerja (506) | Values Bekerja (506), Kuliah (475) |
| ▼ Nilai Akademik | Real | 0 | Min 72 | Max 88 | Average 79.222 |
| ▼ Nilai Non Akademik | Real | 0 | Min 72.060 | Max 88 | Average 79.151 |

Gambar 4. 6 *Dataset Sesudah Missing Value*

| Row No. | NAMA ↑ | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|----------------|---------------|---------|--------------|----------------|-----------------|
| 321 | ROFIIF SYAM... | laki | IPA | Bekerja | 72.670 | 79.151 |

Gambar 4.7 *Dataset yang missing value-nya sudah di replace*

c. Duplikat Data

Duplikat data adalah entri yang sama muncul lebih dari sekali dalam kumpulan data. Ini dapat menyebabkan masalah seperti pemborosan ruang dan ketidakakuratan analisis.

| Row No. | NAMA | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|-------------------|---------------|---------|--------------|----------------|-----------------|
| 1 | Abe1 Tri Yunita | laki | IPA | Bekerja | 77.330 | 77.200 |
| 2 | Ade1ia Dwi 0... | perempuan | IPA | Bekerja | 77.800 | 78.070 |
| 3 | Adi1ia Dwi Wi... | perempuan | IPA | Bekerja | 80.270 | 80.130 |
| 4 | Adinda Intani... | perempuan | IPA | Bekerja | 77.870 | 77.600 |
| 5 | Aditya Dwi Ma... | laki | IPA | Kuliah | 76.730 | 77.200 |
| 6 | Afifah Taskia | perempuan | IPA | Bekerja | 83.470 | 83.070 |
| 7 | Ages Se0rina | laki | IPA | Kuliah | 81.470 | 81.070 |
| 8 | Ajeng Fibri Or... | perempuan | IPA | Bekerja | 81 | 80.670 |
| 9 | Akbar Kurnia... | laki | IPS | Kuliah | 78.330 | 77.730 |
| 10 | A1 Kurniati | perempuan | IPA | Kuliah | 78.670 | 78.600 |
| 11 | A1amsyah | laki | IPA | Kuliah | 80.330 | 80.130 |
| 12 | A1esya A1 Fa... | perempuan | IPA | Bekerja | 79.330 | 79.800 |
| 13 | A1fina Dwi Fe... | perempuan | IPS | Kuliah | 78.800 | 79.130 |

Gambar 4.8 *Dataset* Sebelum *Remove* Duplikat

Dataset ini tidak ditemukan duplikat data. Untuk menghapus duplikat, digunakan fitur "*Remove Duplicates*". Setelah proses tersebut, jumlah *dataset* yang semula berjumlah 981 menjadi 981 tetap sama. Jumlah data yang telah dihapus duplikatnya dapat dilihat lebih jelas pada Gambar 4.9

| Row No. | NAMA | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|------------------|---------------|---------|--------------|----------------|-----------------|
| 1 | Abe1 Tri Yunita | laki | IPA | Bekerja | 77.330 | 77.200 |
| 2 | Ade1ia Dwi 0... | perempuan | IPA | Bekerja | 77.800 | 78.070 |
| 3 | Adi1ia Dwi Wi... | perempuan | IPA | Bekerja | 80.270 | 80.130 |
| 4 | Adinda Intani... | perempuan | IPA | Bekerja | 77.870 | 77.600 |
| 5 | Aditya Dwi Ma... | laki | IPA | Kuliah | 76.730 | 77.200 |
| 6 | Afifah Taskia | perempuan | IPA | Bekerja | 83.470 | 83.070 |
| 7 | Ages Se0rina | laki | IPA | Kuliah | 81.470 | 81.070 |

Gambar 4.9 *Dataset* Sesudah *Remove* Duplikat

d. *Select Attributes*

Select Attributes digunakan untuk memilih kolom - kolom mana saja yang akan digunakan sebagai acuan kluster, pada penelitian kali ini kolom yang akan digunakan antara lain adalah Jenis Kelamin, Jurusan Status Alumni, Nilai

Akademik dan Nilai Non Akademik karena kolom nama terlalu banyak perbedaan sehingga tidak dapat digunakan untuk klusterisasi. Untuk lebih jelasnya dapat dilihat pada gambar 4.10 terkait data yang kolomnya belum di pilih. Sedangkan untuk melihat hasil dari fungsi *Select Attributes* dapat dilihat pada gambar 4.11.

Open in Turbo Prep Auto Model Interactive Analysis

| Row No. | NAMA | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|-------------------|---------------|---------|--------------|----------------|-----------------|
| 1 | Abe1 Tri Yunita | laki | IPA | Bekerja | 77.330 | 77.200 |
| 2 | Ade1ia Dwi 0... | perempuan | IPA | Bekerja | 77.800 | 78.070 |
| 3 | Adi1ia Dwi Wi... | perempuan | IPA | Bekerja | 80.270 | 80.130 |
| 4 | Adinda Intani... | perempuan | IPA | Bekerja | 77.870 | 77.600 |
| 5 | Aditya Dwi Ma... | laki | IPA | Kuliah | 76.730 | 77.200 |
| 6 | Affah Taskia | perempuan | IPA | Bekerja | 83.470 | 83.070 |
| 7 | Ages Se0rina | laki | IPA | Kuliah | 81.470 | 81.070 |
| 8 | Ajeng Fibri Or... | perempuan | IPA | Bekerja | 81 | 80.670 |

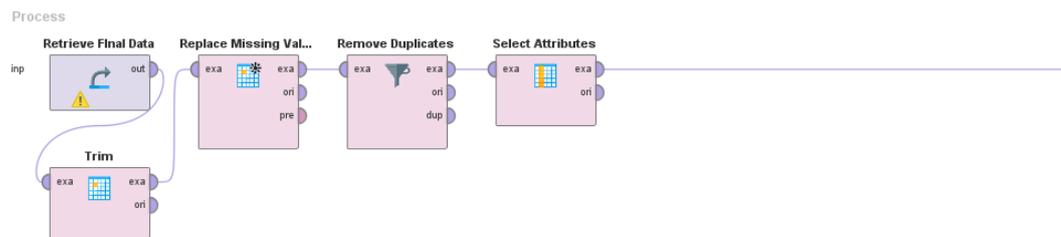
Gambar 4.10 Kolom *Dataset* sebelum dilakukan *Select Attributes*

| Row No. | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|---------------|---------|--------------|----------------|-----------------|
| 1 | laki | IPA | Bekerja | 77.330 | 77.200 |
| 2 | perempuan | IPA | Bekerja | 77.800 | 78.070 |
| 3 | perempuan | IPA | Bekerja | 80.270 | 80.130 |
| 4 | perempuan | IPA | Bekerja | 77.870 | 77.600 |
| 5 | laki | IPA | Kuliah | 76.730 | 77.200 |
| 6 | perempuan | IPA | Bekerja | 83.470 | 83.070 |
| 7 | laki | IPA | Kuliah | 81.470 | 81.070 |
| 8 | perempuan | IPA | Bekerja | 81 | 80.670 |
| 9 | laki | IPS | Kuliah | 78.330 | 77.730 |
| 10 | perempuan | IPA | Kuliah | 78.670 | 78.600 |
| 11 | laki | IPA | Kuliah | 80.330 | 80.130 |
| 12 | perempuan | IPA | Bekerja | 79.330 | 79.800 |
| 13 | perempuan | IPS | Kuliah | 78.800 | 79.130 |

Gambar 4.11 Kolom *Dataset* setelah dilakukan *Select Attributes*

e. Proses

Proses *Cleaning Data* ini dilakukan menggunakan fitur “*Trim*”, “*Replace Missing Value*”, “*Remove Duplicates*” dan “*Select Attributes*” di *RapidMiner*. Tahap pertama, nilai yang hilang digantikan dengan nilai pengganti yang sesuai, sementara pada tahap kedua, data duplikat dihapus untuk menghasilkan *dataset* yang lebih bersih dan efisien. Proses ini dapat dilihat lebih jelas pada Gambar 4.12

Gambar 4.12 Tahapan *Cleaning*

4.2.3 Transformation (Numerical)

Transformasi data *numerical* adalah proses mengubah data *numerical* agar lebih siap untuk analisis atau pemodelan. Teknik yang umum digunakan termasuk normalisasi untuk mengubah rentang data, standarisasi untuk menghasilkan rata-rata 0 dan deviasi standar 1, serta *log transformation* untuk mengurangi dampak nilai ekstrem. Transformasi ini bertujuan meningkatkan kualitas data dan kinerja model.

Tabel 4.2 *Dataset* Sebelum Transformasi

| Jenis Kelamin | Jurusan | Nilai Akademik | Nilai Non Akademik | Status Alumni |
|---------------|---------|----------------|--------------------|---------------|
| laki | ipa | 77.33 | 77.20 | Bekerja |
| perempuan | ipa | 77.80 | 78.07 | Bekerja |
| perempuan | ipa | 80.27 | 80.13 | Bekerja |
| perempuan | ipa | 77.87 | 77.60 | Bekerja |
| laki | ipa | 76.73 | 77.20 | Kuliah |
| perempuan | ipa | 83.47 | 83.07 | Bekerja |
| laki | ipa | 81.47 | 81.07 | Kuliah |

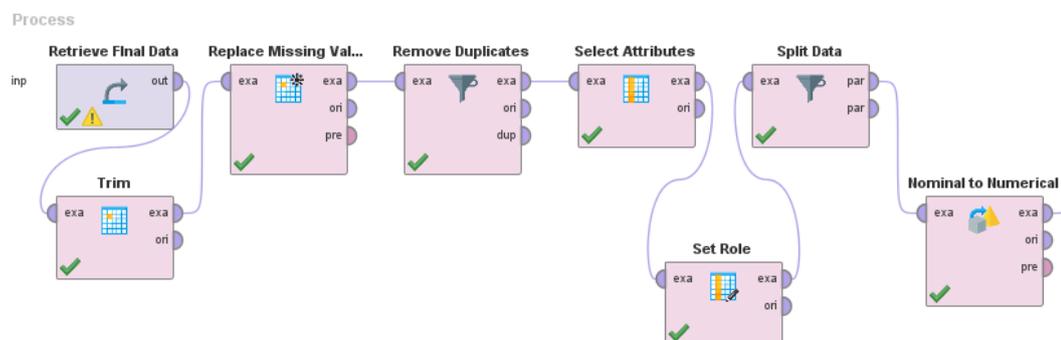
Dataset ini di Transformasi data yang perlu diubah pada atribut Jenis Kelamin, Jurusan dan Status alumni. *Dataset* yang awalnya katagorikal menjadi *dataset numerical* agar *dataset* mudah dianalisis atau dimodelkan. Bisa dilihat pada Gambar 4.13

| Row No. | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|---------------|---------|--------------|----------------|-----------------|
| 1 | 0 | 0 | 0 | 77.330 | 77.200 |
| 2 | 1 | 0 | 0 | 77.800 | 78.070 |
| 3 | 1 | 0 | 0 | 80.270 | 80.130 |
| 4 | 1 | 0 | 0 | 77.870 | 77.600 |
| 5 | 0 | 0 | 1 | 76.730 | 77.200 |
| 6 | 1 | 0 | 0 | 83.470 | 83.070 |
| 7 | 0 | 0 | 1 | 81.470 | 81.070 |

Gambar 4. 13 *Dataset* Sesudah Transformasi

4.2.4 Proses Transpormasi Data (*numerical*)

Proses transformasi data (*numerical*), data numerik ditransformasikan menggunakan teknik seperti normalisasi, standarisasi, *log transformation*, atau *scaling* untuk meningkatkan kualitas data dan kinerja model. Proses ini dapat dilihat pada Gambar 4.14



Gambar 4.14 Tahapan Transformasi Data

4.3 Data Mining

Tahap *Data Mining* data akan diolah dan dianalisis untuk menemukan pola atau informasi penting menggunakan teknik *clustering*. Teknik ini menggunakan algoritma *K-Means*, yang bertujuan untuk mengelompokkan data berdasarkan kesamaan tertentu. Proses pengolahan dan analisis data ini dilakukan menggunakan *RapidMiner* untuk memastikan hasil yang lebih akurat dan efisien.

4.4 Implementasi *K-Means Clustering*

Tahap *Data Mining* akan dilakukan proses pengolahan data dan pencarian pola atau informasi dengan menggunakan teknik *Clustering* dengan menggunakan Aplikasi *RapidMiner*, yaitu proses pengelompokkan data berdasarkan data akademik untuk mengetahui pola nilai dari data akademik sehingga didapatkan informasi dan memudahkan pihak SMA Negeri 1 Pagelaran dalam menentukan strategi meningkatkan mutu sekolah secara tepat.

a. *Import Data*

Import data dengan format *.xlsx* terlebih dahulu ke dalam *Rapidminer* dengan klik "*Import Data*". Bisa dilihat pada Gambar 4.15



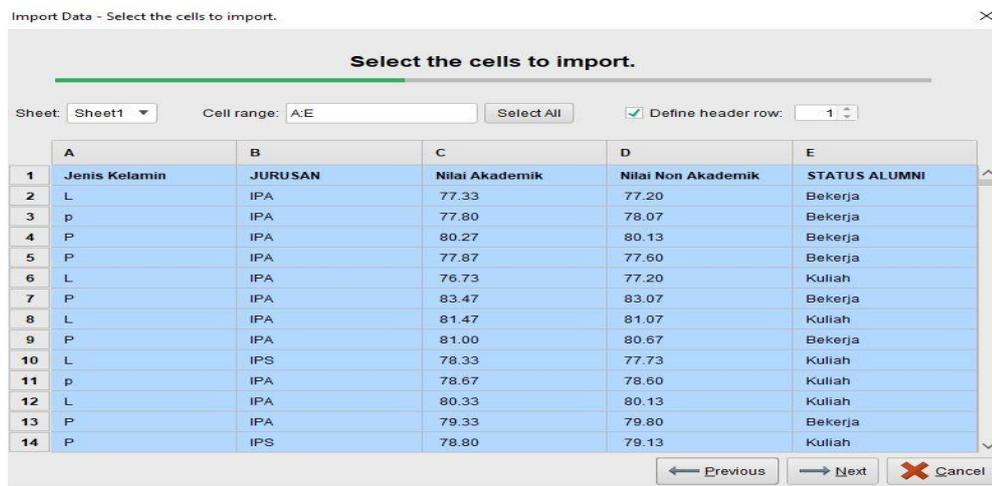
Gambar 4.15 *Import Data*

Data yang sudah diimport kemudian pilih "*My Computer*" untuk memilih *dataset* yang akan digunakan. Bisa dilihat pada Gambar 4.16



Gambar 4.16 Input Data

Berdasarkan *dataset* yang sudah berhasil diimpor pada *Rapidminer dataset* akan muncul seperti Gambar 4.17



Gambar 4.17 Data Yang Berhasil Diimpor

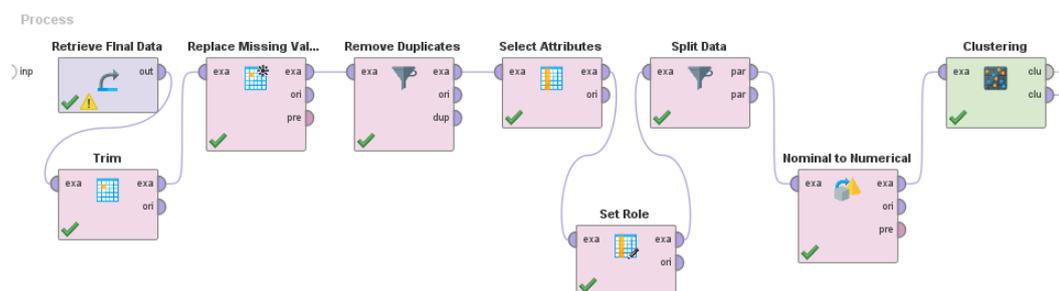
b. Tahapan Pemodelan *Cluster*

Teknik *clustering* adalah pendekatan dalam analisis data yang bertujuan untuk mengelompokkan entitas data secara otomatis berdasarkan kesamaan karakteristiknya, tanpa adanya label kelas yang telah ditentukan sebelumnya. *Clustering* yang sering dikategorikan sebagai metode *unsupervised learning*, memungkinkan klasifikasi data yang belum memiliki label kelas yang jelas atau terdefinisi. Metode ini sangat penting dalam analisis data yang kompleks, terutama ketika informasi mengenai struktur kelas tidak diketahui atau sulit ditetapkan. Proses *clustering* berfokus pada pengenalan pola-pola intrinsik dalam data untuk

membentuk kelompok-kelompok homogen dengan tingkat kemiripan tinggi di antara anggotanya.

Teknik *clustering* adalah pendekatan dalam analisis data yang bertujuan untuk mengelompokkan entitas data secara otomatis berdasarkan kesamaan karakteristiknya, tanpa adanya label kelas yang telah ditentukan sebelumnya. *Clustering* yang sering dikategorikan sebagai metode *unsupervised learning*, memungkinkan klasifikasi data yang belum memiliki label kelas yang jelas atau terdefinisi. Metode ini sangat penting dalam analisis data yang kompleks, terutama ketika informasi mengenai struktur kelas tidak diketahui atau sulit ditetapkan. Proses *clustering* berfokus pada pengenalan pola-pola intrinsik dalam data untuk membentuk kelompok-kelompok homogen dengan tingkat kemiripan tinggi di antara anggotanya.

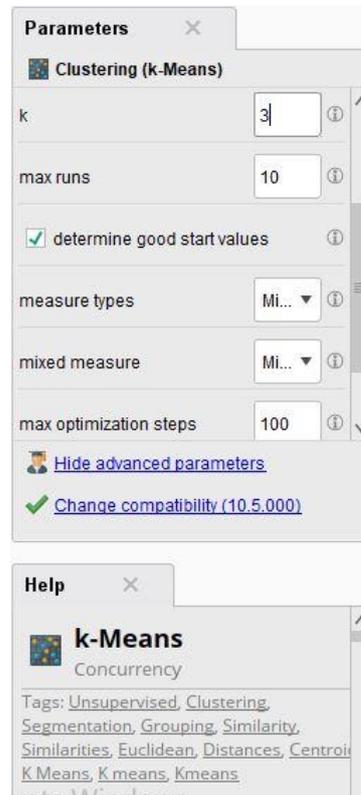
Penelitian ini, saya menerapkan metode klasterisasi *k-means*. Algoritma ini beroperasi dengan menentukan jumlah *Cluster* (k) yang diinginkan. Selain itu, *k-means* bekerja dengan menunggu hingga jumlah maksimum iterasi tercapai atau hingga memenuhi nilai konstanta tertentu yang digunakan sebagai kriteria penghentian iterasi. Bisa dilihat pada Gambar 4.18



Gambar 4.18 Tahapan Pemodelan *Cluster*

Parameter algoritma *clustering*, saya melakukan simulasi sebanyak tujuh kali dalam pembentukan kelas untuk menentukan model yang paling optimal. Sementara itu, nilai *max run* disesuaikan dengan nilai *default*, yaitu 10 kali, guna mengevaluasi konsistensi model. Pada *measures type*, saya memilih *mixed measures* karena data dalam data set bersifat *heterogen*. Bisa dilihat pada Gambar 4.19 untuk pengaturan

parameter algoritma *k-means* dalam perhitungan data set Akademik.



Gambar 4.19 Parameter *Cluster*

Hasil dari pembentukan *Cluster* dari nilai $k=2$ sampai $k=7$ sebagai berikut:

Tabel 4.3 Penjelasan Pembentukan *Cluster* 1 hingga 7

| | |
|-----|---|
| K-2 | <p>Hasil dari model <i>klastering</i> pada <i>RapidMiner</i> menghasilkan dua <i>klaster</i> utama. <i>Cluster</i> 0, dengan jumlah 426 item, dan <i>Cluster</i> 1, dengan jumlah 555 item, memberikan gambaran mengenai pemisahan data menjadi dua kelompok yang signifikan. <i>Cluster</i> ini mencerminkan pola-pola atau karakteristik tertentu yang membedakan antara dua kelompok tersebut. <i>Clusetr</i> dengan jumlah item terendah adalah <i>Cluster</i> 0, sedangkan <i>Cluster</i> 1 memiliki jumlah item tertinggi. Perbedaan jumlah item ini mengindikasikan bahwa <i>Cluster</i> 1 mungkin memiliki variasi atau kompleksitas yang lebih tinggi dibandingkan dengan <i>Cluster</i> 0</p> |
|-----|---|

| | |
|-----|--|
| K-3 | <p><i>Cluster 1</i> dengan 223 item, dan <i>Cluster 2</i> dengan 296 item. Analisis lebih lanjut mengungkapkan bahwa <i>Cluster 2</i> merupakan <i>Cluster</i> tertinggi dengan jumlah item terbanyak, sedangkan <i>Cluster 1</i> adalah <i>Cluster</i> dengan jumlah item terendah</p> |
| K-4 | <p><i>Cluster 0</i> menjadi <i>Cluster</i> dengan jumlah data terbanyak, mencapai 296 item, mengindikasikan sekelompok besar data yang memiliki kesamaan karakteristik tertentu. <i>Cluster</i> dengan jumlah data terendah adalah <i>Cluster 2</i>, yang terdiri dari 120 item. Setiap <i>Cluster</i> mewakili kelompok data yang memiliki tingkat kemiripan yang tinggi di antara anggotanya, sementara perbedaan karakteristik antar <i>Cluster</i> dapat diidentifikasi.</p> |
| K-5 | <p><i>Cluster 0</i> memiliki sebanyak 348 entitas data, <i>Cluster 1</i> terdiri dari 106 entitas data, sementara <i>Cluster 2</i> mencakup 202 entitas data. <i>Cluster 3</i> memiliki 284 entitas data, dan <i>Klaster 4</i> terdiri dari 41 entitas data.</p> |
| K-6 | <p><i>Cluster 0</i> memiliki 260 item, <i>Cluster 1</i> mencakup 41 item, sementara <i>Cluster 2</i> terdiri dari 81 item. <i>Cluster 3</i> memiliki 172 item, <i>Cluster 4</i> terdiri dari 206 item, dan <i>Cluster 5</i> mencakup 221 item</p> |
| K-7 | <p><i>Cluster 0</i> terdiri dari 81 item, sementara <i>Cluster 1</i> memiliki 164 item. Selanjutnya, <i>Cluster 2</i> mencakup 259 item, <i>Cluster 3</i> terdiri dari 206 item, dan <i>Cluster 4</i> memiliki 41 item. <i>Cluster 5</i> terdiri dari 220 item, <i>Cluster 6</i> terdiri dari 10 item.</p> |

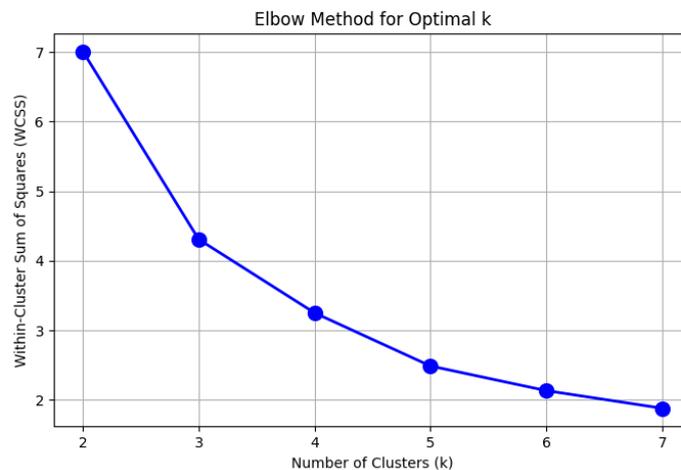
c. Tahapan Analisa *Cluster*

Analisa *Cluster* adalah salah satu teknik dalam *data mining* yang digunakan untuk mengidentifikasi *Cluster* dalam himpunan data secara otomatis atau semi-otomatis. Berdasarkan pendekatan dan konsep yang diterapkan, teknik *clustering* dapat diklasifikasikan ke dalam metode berbasis partisi, hierarki, densitas/kerapatan, *grid*, model, dan *constraint*. Dalam penelitian ini, evaluasi kinerja model *clustering* dilakukan menggunakan metode *Within-Cluster Sum of Squares* (WCSS). WCSS digunakan untuk mengukur total jarak kuadrat antara setiap titik data dalam *cluster* dengan *centroid*-nya. Semakin kecil nilai WCSS, semakin baik kualitas *cluster* yang terbentuk karena menunjukkan bahwa data dalam *cluster* lebih homogen dan berdekatan dengan *centroid*-nya. Pada penelitian ini, saya melakukan simulasi model untuk menentukan *Cluster* terbaik pada data set akademik. Bisa dilihat pada Tabel 4.4 untuk hasil simulasi penentuan *Cluster* menggunakan *Within Cluster Sum of Squares* (WCSS):

Tabel 4. 4 Hasil Analisa *Cluster* Perhitungan WCSS

| Banyak Klaster (K) | Hasil |
|--------------------|-------|
| 2 | 7.011 |
| 3 | 4.308 |
| 4 | 3.251 |
| 5 | 2.487 |
| 6 | 2.133 |
| 7 | 1.879 |

Untuk menentukan nilai *Cluster* (k) terbaik pada data set akademik, saya menggunakan metode *elbow*. Metode ini bertujuan untuk menentukan jumlah *Cluster* yang optimal sehingga hasil *klasterisasi* sesuai dengan struktur intrinsik data, tanpa mengalami *overfitting* maupun *underfitting*. Bisa dilihat pada Gambar 4.20 untuk penerapan metode *elbow*



Gambar 4.20 *Elbow Method*

Grafik *Elbow* menunjukkan hubungan antara jumlah *cluster* (k) dengan nilai WCSS. Pada grafik ini, sumbu horizontal (X) menunjukkan jumlah *cluster* (k), sementara sumbu vertikal (Y) menunjukkan nilai WCSS. Penurunan nilai WCSS yang signifikan terjadi hingga $k = 3$. Setelah jumlah *cluster* lebih dari 3, laju penurunan WCSS mulai melambat. Titik di mana penurunan ini mulai melambat disebut "titik *elbow*," yang mengindikasikan jumlah *cluster* optimal. Dari grafik tersebut, terlihat bahwa titik *elbow* berada pada $k = 3$, yang berarti jumlah *cluster* optimal adalah 3.

d. Tahapan Kesimpulan

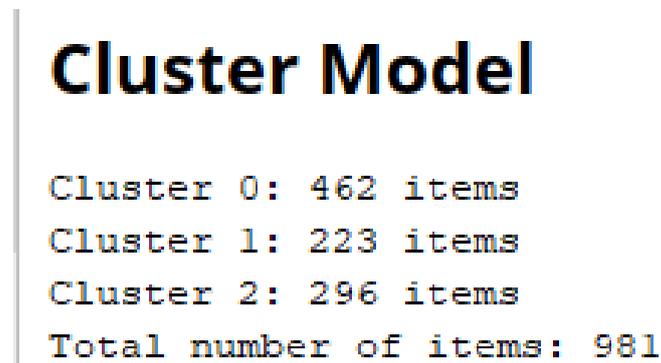
Berdasarkan Analisa di atas, dapat disimpulkan beberapa hasil penelitian sebagai berikut:

1. Berdasarkan Analisa data set Akademik memiliki nilai *Cluster* terbaik $k = 3$. Nilai rata-rata jarak antar *centroid* adalah 4.308. nilai rata-rata jarak antar *centroid*: 4.308, dapat diartikan bahwa rata-rata jarak antara titik data dalam *Cluster* dengan pusat *Cluster* adalah sekitar 4.308. Semakin kecil nilai ini, semakin padat dan *homogen Cluster* tersebut. Bisa dilihat pada Gambar 4.21



Gambar 4.21 Rata-rata Dalam Jarak Centroid

2. Rincian masing-masing *Cluster* dibagi menjadi *Cluster 0* dengan jumlah 462 item, *Cluster 1* dengan jumlah 223 item dan *Cluster 2* dengan jumlah 296 item. Bisa dilihat pada Gambar 4.22 untuk *cluster* model yang terbentuk.



Gambar 4. 22 Model Klaster

3. Interpretasi Hasil *Clustering*, hasil *clustering* dianalisis berdasarkan karakteristik utama dari setiap *cluster*. Berikut adalah hasil evaluasi *cluster* berdasarkan *dataset* akademik SMA Negeri 1 Pagelaran.
 - a. Interpretasi Berdasarkan Jenis Kelamin
 - **Cluster 0:** Memiliki 178 siswa laki-laki dan 284 siswa perempuan.
 - **Cluster 1:** Memiliki 77 siswa laki-laki dan 146 siswa perempuan.
 - **Cluster 2:** Memiliki 101 siswa laki-laki dan 195 siswa perempuan.
 - b. Interpretasi Berdasarkan Jurusan
 - **Cluster 0:** Terdiri dari 272 siswa IPA dan 190 siswa IPS.
 - **Cluster 1:** Terdiri dari 139 siswa IPA dan 84 siswa IPS.
 - **Cluster 2:** Terdiri dari 182 siswa IPA dan 114 siswa IPS.
 - c. Interpretasi Berdasarkan Status Alumni

- **Cluster 0:** Terdiri dari 243 siswa yang bekerja dan 219 siswa yang melanjutkan kuliah.
 - **Cluster 1:** Terdiri dari 110 siswa yang bekerja dan 113 siswa yang melanjutkan kuliah.
 - **Cluster 2:** Terdiri dari 153 siswa yang bekerja dan 143 siswa yang melanjutkan kuliah.
- d. Interpretasi Berdasarkan Nilai Akademik
- **Cluster 0:** Siswa dengan nilai akademik sedang.
 - **Cluster 1:** Siswa dengan nilai akademik tinggi.
 - **Cluster 2:** Siswa dengan nilai akademik rendah.
- e. Interpretasi Berdasarkan Nilai Non-Akademik
- **Cluster 0:** Siswa dengan nilai non-akademik sedang.
 - **Cluster 1:** Siswa dengan nilai non-akademik tinggi.
 - **Cluster 2:** Siswa dengan nilai non-akademik rendah.

Dari analisis ini, sekolah dapat menggunakan hasil *clustering* untuk menyusun strategi peningkatan mutu pendidikan, seperti memberikan perhatian khusus kepada siswa dalam *Cluster 2* yang memiliki nilai rendah atau mengembangkan program peningkatan kompetensi untuk siswa di *Cluster 0*.

Berdasarkan hasil evaluasi menggunakan WCSS dan metode *Elbow*, model *clustering* dengan $k = 3$ menunjukkan performa terbaik dalam mengelompokkan data akademik siswa. Nilai WCSS yang lebih rendah menunjukkan bahwa pengelompokan data memiliki tingkat homogenitas yang baik dalam setiap *cluster*. Dengan informasi ini, sekolah dapat menerapkan strategi pembelajaran yang lebih spesifik dan efektif sesuai dengan kebutuhan masing-masing kelompok siswa.

4.6 Hasil Penelitian

Berdasarkan hasil analisis data dengan metode *k-means*, diperoleh bahwa jumlah *Cluster* yang optimal adalah ketika data dibagi menjadi tiga *Cluster* bisa dilihat pada Gambar 4.23

| Row No. | id | cluster | Jenis Kelamin | JURUSAN | STATUS AL... | Nilai Akade... | Nilai Non Ak... |
|---------|----|-----------|---------------|---------|--------------|----------------|-----------------|
| 1 | 1 | cluster_2 | 0 | 0 | 0 | 77.330 | 77.200 |
| 2 | 2 | cluster_0 | 1 | 0 | 0 | 77.800 | 78.070 |
| 3 | 3 | cluster_0 | 1 | 0 | 0 | 80.270 | 80.130 |
| 4 | 4 | cluster_2 | 1 | 0 | 0 | 77.870 | 77.600 |
| 5 | 5 | cluster_2 | 0 | 0 | 1 | 76.730 | 77.200 |
| 6 | 6 | cluster_1 | 1 | 0 | 0 | 83.470 | 83.070 |
| 7 | 7 | cluster_1 | 0 | 0 | 1 | 81.470 | 81.070 |
| 8 | 8 | cluster_0 | 1 | 0 | 0 | 81 | 80.670 |
| 9 | 9 | cluster_0 | 0 | 1 | 1 | 78.330 | 77.730 |
| 10 | 10 | cluster_0 | 1 | 0 | 1 | 78.670 | 78.600 |
| 11 | 11 | cluster_0 | 0 | 0 | 1 | 80.330 | 80.130 |
| 12 | 12 | cluster_0 | 1 | 0 | 0 | 79.330 | 79.800 |
| 13 | 13 | cluster_0 | 1 | 1 | 1 | 78.800 | 79.130 |
| 14 | 14 | cluster_0 | 0 | 0 | 0 | 79.870 | 79.930 |
| 15 | 15 | cluster_0 | 1 | 1 | 1 | 79.070 | 79 |
| 16 | 16 | cluster_0 | 0 | 0 | 1 | 78.670 | 78.470 |

Gambar 4. 23 Hasil Penelitian

Berdasarkan gambar 4.24 menunjukkan hasil analisis statistik yang mendukung temuan penelitian. Visualisasi ini mempermudah pemahaman data.

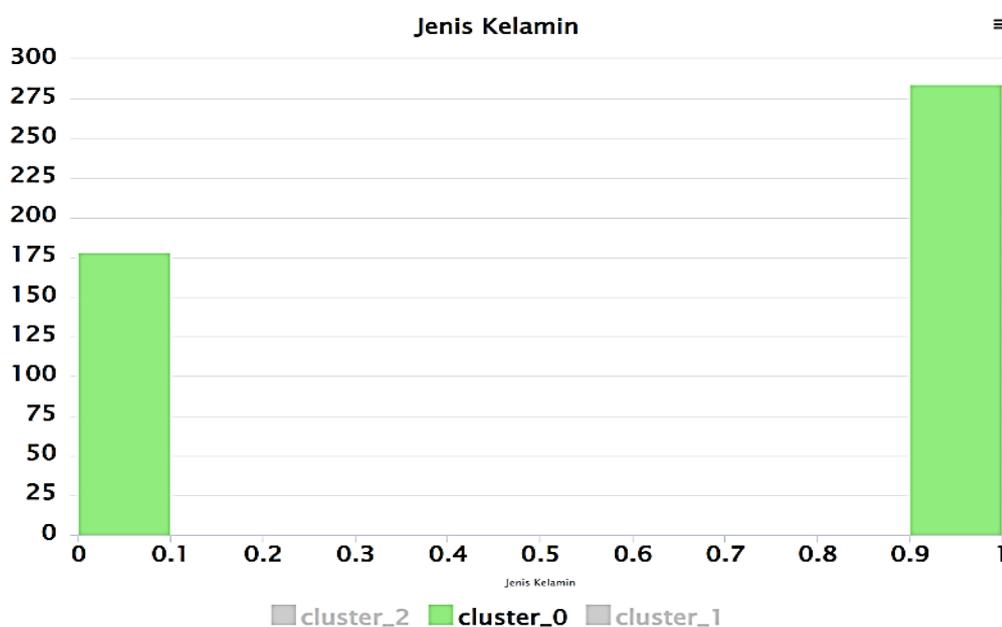


Gambar 4. 24 Statistik pada Rapidminer

Berdasarkan hasil klasterisasi, maka data set akademik dengan jumlah 981 dibagi menjadi 3 karakteristik kemiripan data. Sebanyak 462 memiliki karakteristik data yang sama dan termasuk dalam *Cluster 0*, Sedangkan 223 data memiliki karakteristik yang mirip serta termasuk dalam *Cluster 1* dan 296 data memiliki karakteristik yang mirip serta termasuk dalam *Cluster 2*. Untuk menentukan karakteristik *Cluster 0*, *Cluster 1* dan *Cluster 2* dalam menentukan penggolongan tinggi, sedang dan rendah nilai siswa.

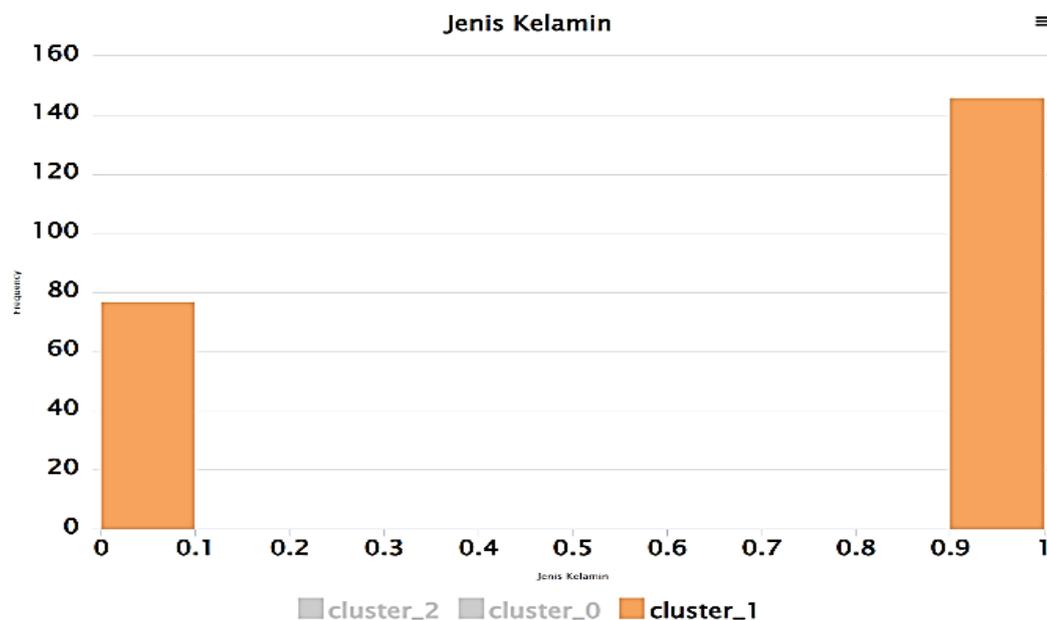
a. Jenis Kelamin

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 0*, terdapat sebanyak 178 siswa laki-laki. Sementara itu, jumlah siswa perempuan yang termasuk dalam *Cluster* yang sama adalah sebanyak 284 siswa. Dari data tersebut, dapat disimpulkan bahwa pada *Cluster 0*, jumlah siswa perempuan lebih banyak dibandingkan dengan jumlah siswa laki-laki. Hal ini menunjukkan bahwa mayoritas anggota dalam *Cluster 0* didominasi oleh siswa laki-laki. Informasi lebih lanjut mengenai distribusi jumlah siswa dalam *Cluster* ini dapat dilihat pada Gambar 4.25



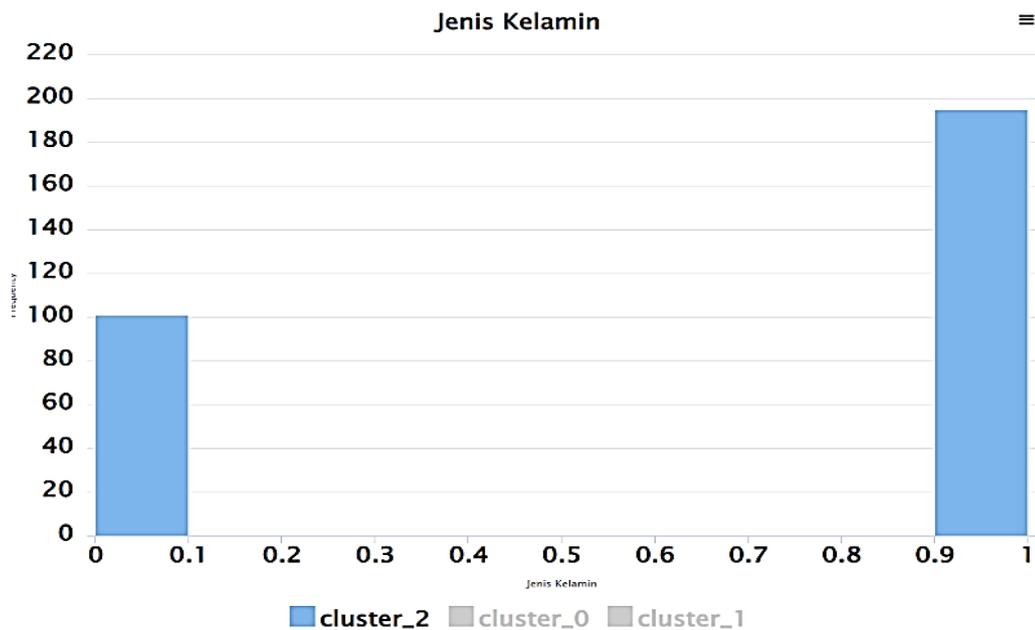
Gambar 4. 25 Visualisasi *Cluster 0* Jenis Kelamin

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 1*, terdapat sebanyak 77 siswa laki-laki yang tergabung dalam kelompok ini. Sementara itu, jumlah siswa perempuan yang termasuk dalam *Cluster* yang sama adalah sebanyak 146 siswa. Dari data tersebut, dapat diketahui bahwa jumlah siswa perempuan dalam *Cluster 1* lebih besar dibandingkan jumlah siswa laki-laki. Hal ini menunjukkan bahwa mayoritas anggota dalam *Cluster 1* didominasi oleh siswa perempuan. Informasi lebih lanjut mengenai distribusi jumlah siswa dalam *Cluster* ini dapat dilihat pada Gambar 4.26



Gambar 4.26 Visualisasi *Cluster 1* Jenis Kelamin

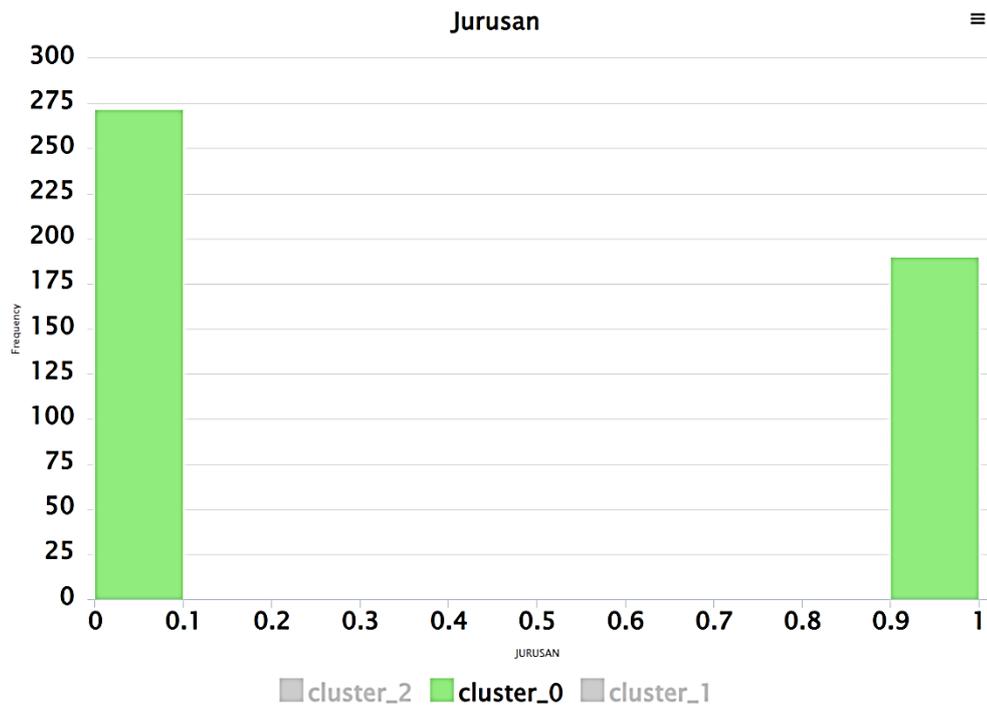
Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 2*, terdapat sebanyak 101 siswa laki-laki yang tergabung dalam kelompok ini. Sementara itu, jumlah siswa perempuan yang termasuk dalam *Cluster* yang sama adalah sebanyak 195 siswa. Dari data tersebut, dapat diketahui bahwa jumlah siswa laki-laki dalam *Cluster 2* lebih besar dibandingkan dengan jumlah siswa perempuan. Dengan demikian, dapat disimpulkan bahwa mayoritas anggota dalam *Cluster 2* didominasi oleh siswa perempuan. Informasi lebih lanjut mengenai distribusi jumlah siswa dalam *Cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.27



Gambar 4. 27 Visualisasi *Cluster 2* Jenis Kelamin

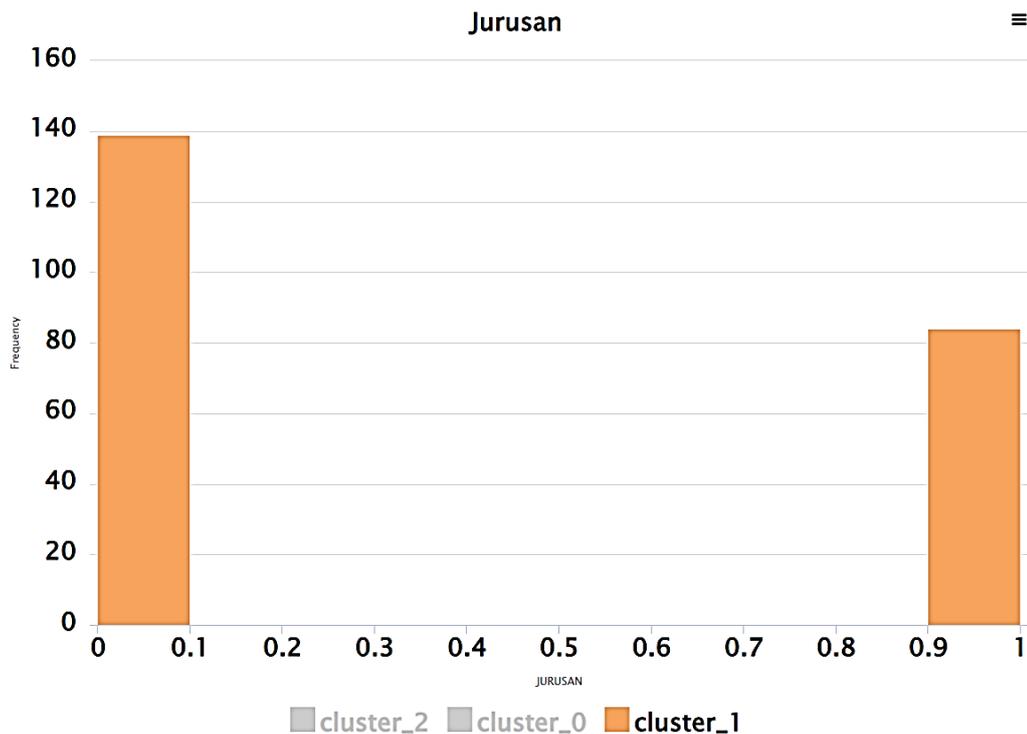
b. Jurusan

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 0*, terdapat sebanyak 272 siswa yang berasal dari jurusan Ilmu Pengetahuan Alam (IPA). Sementara itu, jumlah siswa yang berasal dari jurusan Ilmu Pengetahuan Sosial (IPS) dalam *Cluster* yang sama adalah sebanyak 190 siswa. Dari data tersebut, dapat diketahui bahwa jumlah siswa dari jurusan IPA lebih banyak dibandingkan dengan jumlah siswa dari jurusan IPS dalam *Cluster 0*. Dengan demikian, dapat disimpulkan bahwa mayoritas anggota dalam *Cluster 0* berasal dari jurusan IPA. Informasi lebih lanjut mengenai distribusi jumlah siswa berdasarkan jurusan dalam *Cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.28



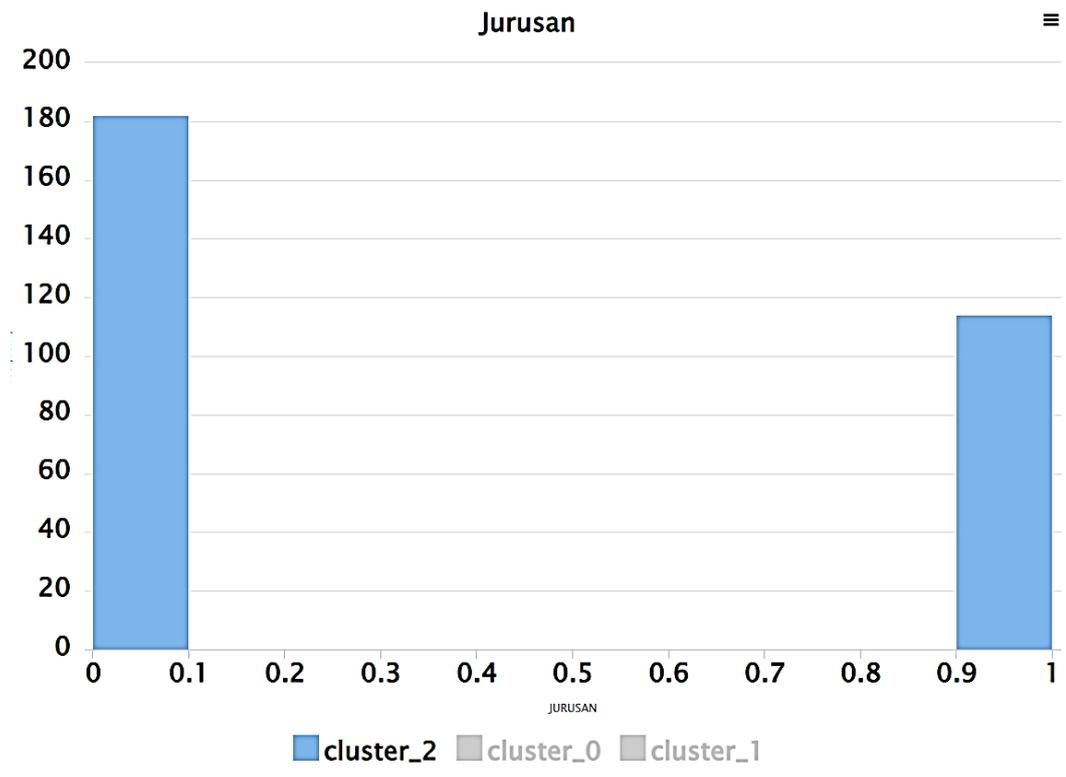
Gambar 4. 28 Visualisasi *Cluster 0* Jurusan

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 1*, terdapat sebanyak 139 siswa yang berasal dari jurusan Ilmu Pengetahuan Alam (IPA). Sementara itu, jumlah siswa yang berasal dari jurusan Ilmu Pengetahuan Sosial (IPS) dalam *Cluster* yang sama adalah sebanyak 84 siswa. Dari data tersebut, dapat diketahui bahwa jumlah siswa dari jurusan IPA lebih banyak dibandingkan dengan jumlah siswa dari jurusan IPS dalam *Cluster 1*. Dengan demikian, dapat disimpulkan bahwa mayoritas anggota dalam *Cluster 1* berasal dari jurusan IPA. Informasi lebih lanjut mengenai distribusi jumlah siswa berdasarkan jurusan dalam *Cluster* ini dapat dilihat pada Gambar 4.29



Gambar 4. 29 Visualisasi *Cluster* 1 Jurusan

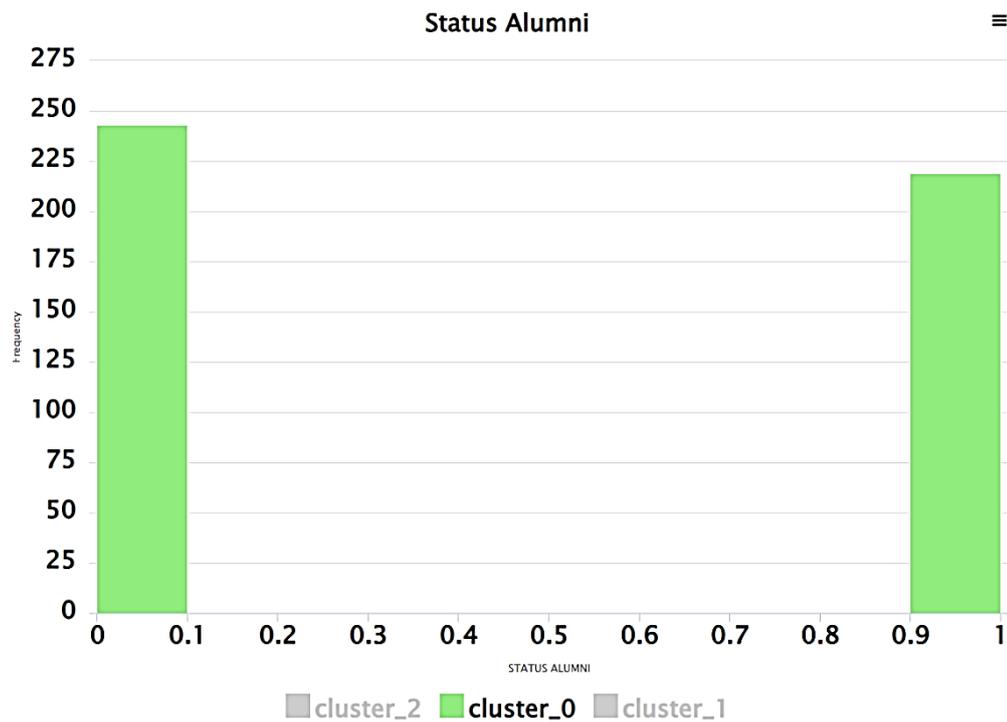
Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 2*, terdapat sebanyak 182 siswa yang berasal dari jurusan Ilmu Pengetahuan Alam (IPA). Sementara itu, jumlah siswa yang berasal dari jurusan Ilmu Pengetahuan Sosial (IPS) dalam *Cluster* yang sama adalah sebanyak 114 siswa. Dari data tersebut, dapat diketahui bahwa jumlah siswa dari jurusan IPA lebih banyak dibandingkan dengan jumlah siswa dari jurusan IPS dalam *Cluster 2*. Dengan demikian, dapat disimpulkan bahwa mayoritas anggota dalam *Cluster 2* berasal dari jurusan IPA. Informasi lebih lanjut mengenai distribusi jumlah siswa berdasarkan jurusan dalam *Cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.30



Gambar 4. 30 Visualisasi *Cluster 2* Jurusan

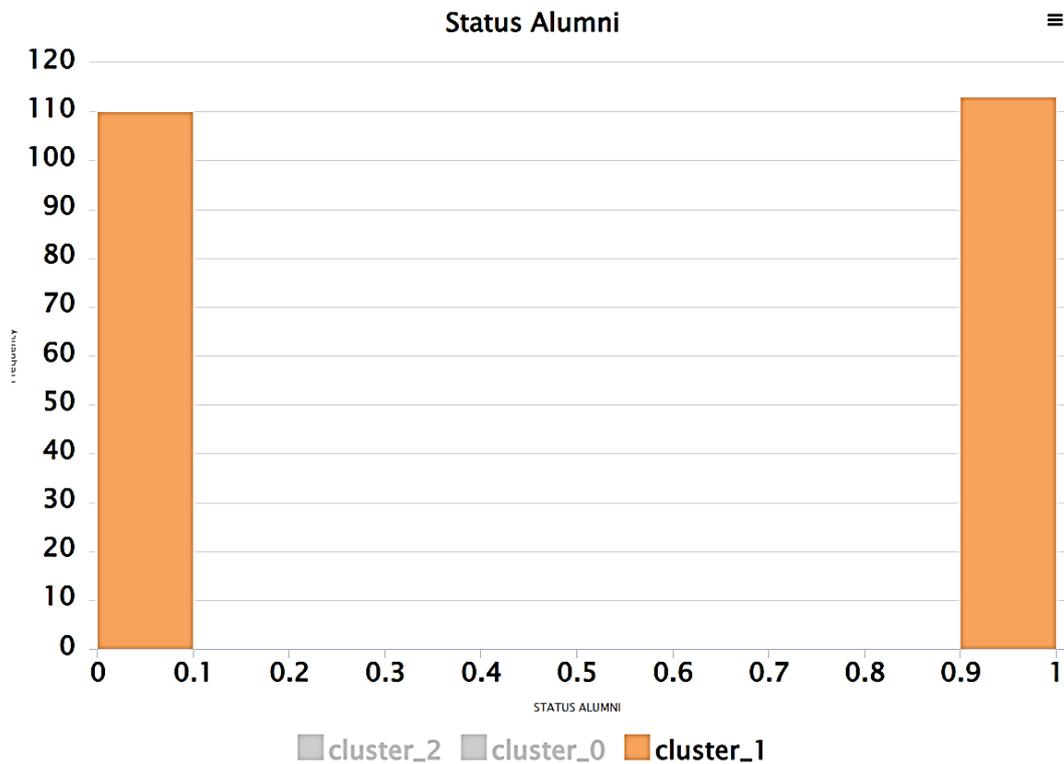
c. Status Alumni

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 0* terdapat sebanyak 243 siswa yang memiliki status alumni sebagai pekerja. Sementara itu, jumlah alumni yang melanjutkan pendidikan ke jenjang perguruan tinggi atau memiliki status kuliah dalam *Cluster* yang sama adalah sebanyak 219 siswa. Dari data tersebut, dapat diketahui bahwa jumlah alumni yang bekerja lebih banyak dibandingkan dengan jumlah alumni yang melanjutkan kuliah dalam *Cluster 0*. Dengan demikian, dapat disimpulkan bahwa mayoritas alumni dalam *Cluster 0* memilih untuk langsung bekerja setelah menyelesaikan pendidikan mereka. Informasi lebih lanjut mengenai distribusi status alumni dalam *Cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.31



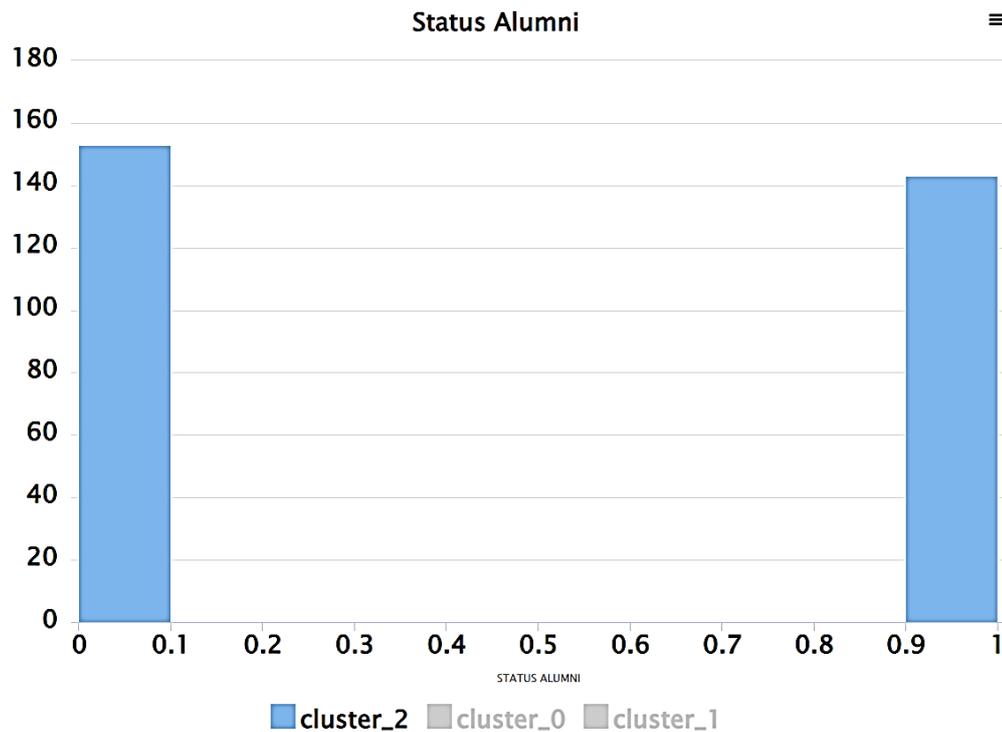
Gambar 4. 31 Visualisasi *Cluster 0* Status Alumni

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 1*, terdapat sebanyak 110 siswa yang memiliki status alumni sebagai pekerja. Sementara itu, jumlah alumni yang melanjutkan pendidikan ke jenjang perguruan tinggi atau memiliki status kuliah dalam *Cluster* yang sama adalah sebanyak 113 siswa. Dari data tersebut, dapat diketahui bahwa jumlah alumni yang memilih untuk kuliah lebih banyak dibandingkan dengan jumlah alumni yang melanjutkan kerja dalam *Cluster 1*. Dengan demikian, dapat disimpulkan bahwa mayoritas alumni dalam *Cluster 1* lebih cenderung masuk kuliah ke perguruan tinggi setelah menyelesaikan pendidikan mereka. Informasi lebih lanjut mengenai distribusi status alumni dalam *Cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.32



Gambar 4. 32 Visualisasi *Cluster* 1 Status Alumni

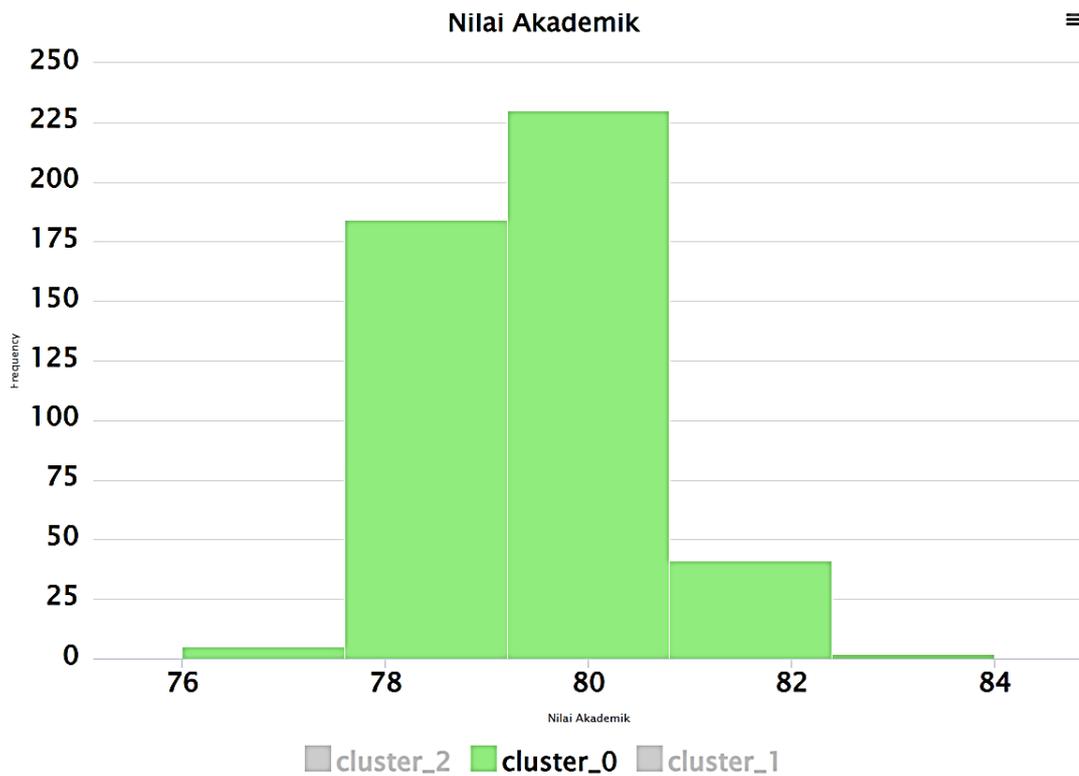
Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *Cluster 2*, terdapat sebanyak 153 siswa yang memiliki status alumni sebagai pekerja. Sementara itu, jumlah alumni yang melanjutkan pendidikan ke jenjang perguruan tinggi atau memiliki status kuliah dalam *Cluster* yang sama adalah sebanyak 143 siswa. Dari data tersebut, dapat diketahui bahwa jumlah alumni yang memilih untuk bekerja lebih banyak dibandingkan dengan jumlah alumni yang melanjutkan kuliah dalam *Cluster 2*. Dengan demikian, dapat disimpulkan bahwa mayoritas alumni dalam *Cluster 2* lebih cenderung memasuki dunia kerja setelah menyelesaikan pendidikan mereka. Informasi lebih lanjut mengenai distribusi status alumni dalam *Cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.33



Gambar 4. 33 Visualisasi *Cluster 2* Status Alumni

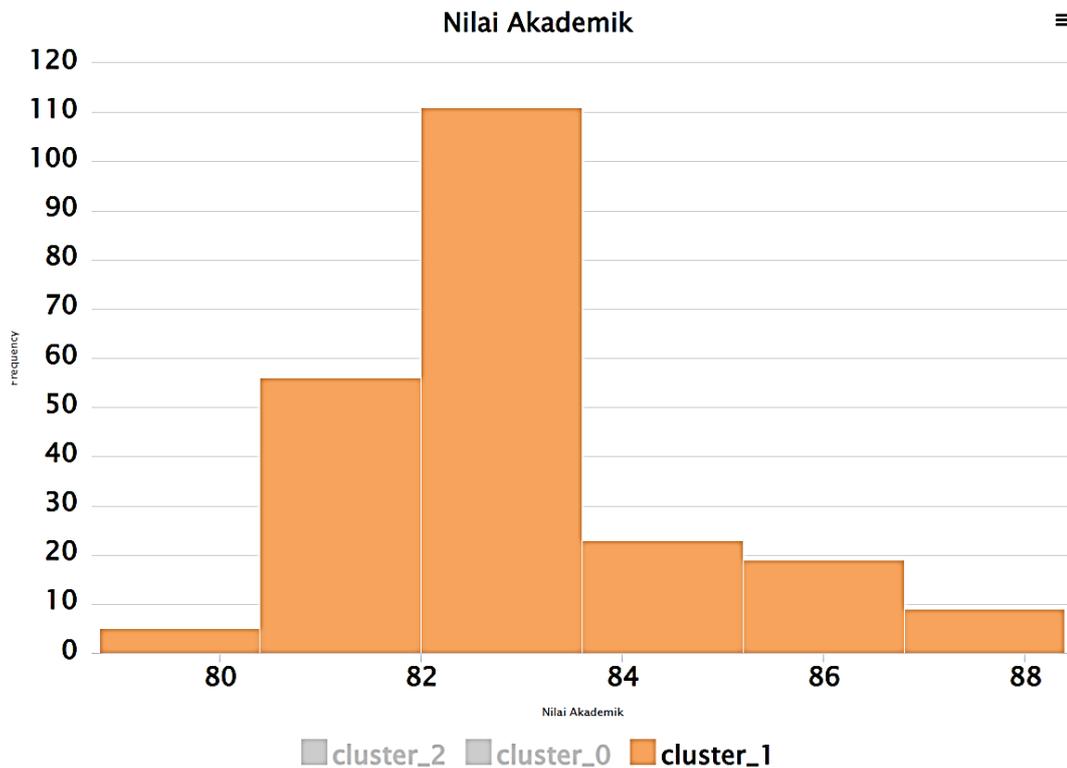
d. Nilai Akademik

Berdasarkan visualisasi hasil klasterisasi untuk kelompok atau *cluster* 0, terdapat berbagai rentang nilai yang diperoleh oleh siswa dengan distribusi sebagai berikut: "Distribusi nilai akademik siswa dalam diagram menunjukkan bahwa terdapat 5 siswa dengan nilai akademik dalam rentang 76-77.6. Sebagian besar siswa, sebanyak 184 orang, memiliki nilai akademik dalam rentang 77.6-79.2. Jumlah siswa tertinggi, yaitu 230 siswa, berada dalam rentang nilai 79.2-80.8. Sementara itu, terdapat 41 siswa yang memiliki nilai akademik dalam rentang 80.8-82.4. Hanya 2 siswa yang memiliki nilai akademik dalam rentang 82.4-84. Data ini menunjukkan bahwa sebagian besar siswa memiliki nilai akademik antara 77.6 hingga 80.8, dengan puncak jumlah siswa berada pada nilai 79.2 hingga 80.8. Hal ini menunjukkan bahwa mayoritas siswa dalam *cluster* ini memiliki nilai yang berkisar dalam rentang tersebut. Informasi lebih lanjut mengenai distribusi nilai dalam *cluster* ini dapat dilihat secara lebih rinci pada Gambar 4.34



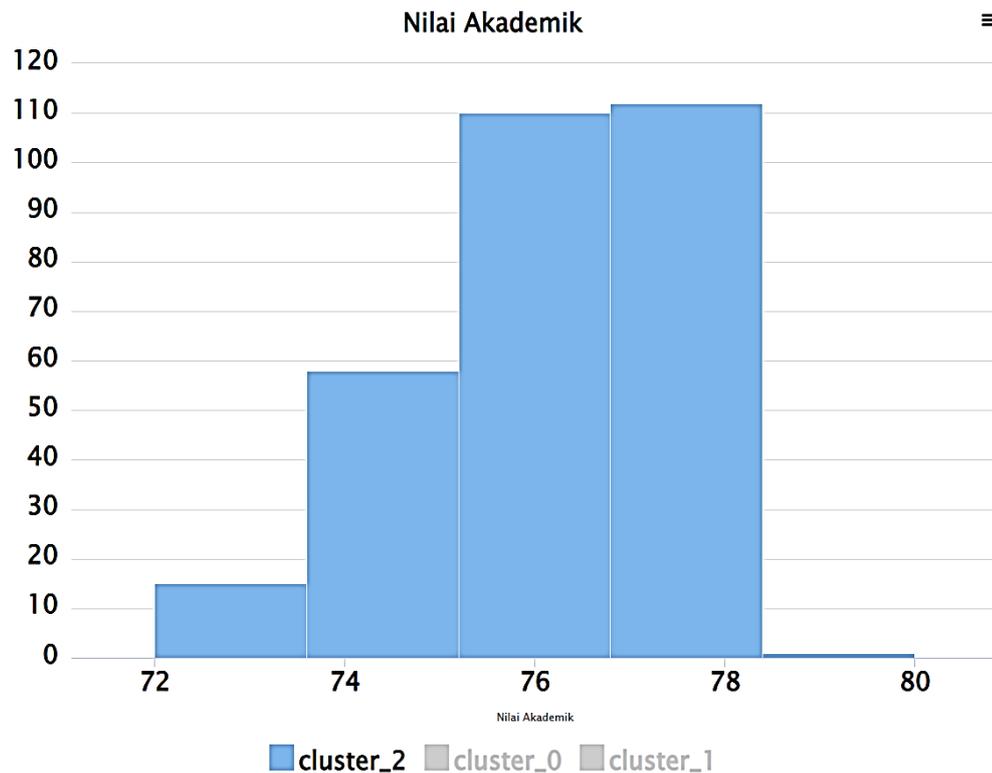
Gambar 4. 34 Visualisasi *Cluster 0* Nilai Akademik

Berdasarkan gambar visualisasi *cluster 1*, distribusi nilai akademik siswa dalam diagram menunjukkan bahwa terdapat 5 siswa dengan nilai akademik dalam rentang 78.8-80.4. Sebanyak 56 siswa memiliki nilai akademik dalam rentang 80.4-82. Jumlah siswa tertinggi, yaitu 111 siswa, berada dalam rentang nilai 82-83.6. Selain itu, terdapat 23 siswa dengan nilai akademik dalam rentang 83.6-85.2, 19 siswa dalam rentang 85.2-86.8, dan 9 siswa dalam rentang 86.8-88.4. Data ini menunjukkan bahwa sebagian besar siswa memiliki nilai akademik antara 80.4 hingga 83.6, dengan puncak jumlah siswa berada pada nilai 82 hingga 83.6. Informasi ini dapat dilihat secara lebih rinci pada Gambar 4.35



Gambar 4. 35 Visualisasi *Cluster 1* Nilai Akademik

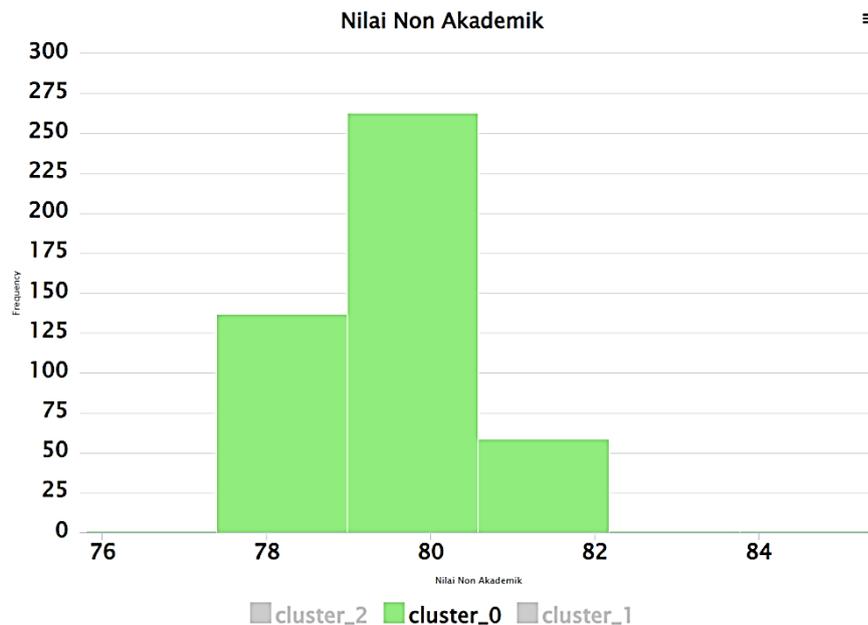
Berdasarkan gambar visualisasi *cluster 2*, distribusi nilai akademik siswa dalam diagram menunjukkan bahwa terdapat 15 siswa dengan nilai akademik dalam rentang 72-73.6. Sebanyak 58 siswa memiliki nilai akademik dalam rentang 73.6-75.2. Jumlah siswa yang signifikan, yaitu 110 siswa, berada dalam rentang nilai 75.2-76.8. Selain itu, jumlah siswa tertinggi, yaitu 112 siswa, memiliki nilai akademik dalam rentang 76.8-78.4. Hanya 1 siswa yang memiliki nilai akademik dalam rentang 78.4-80. Data ini menunjukkan bahwa sebagian besar siswa memiliki nilai akademik antara 72 hingga 78.4, dengan puncak jumlah siswa berada pada nilai 76.8 hingga 78.4. Informasi lebih rinci mengenai distribusi nilai dalam *cluster* ini dapat dilihat pada Gambar 4.36



Gambar 4. 36 Visualisasi *Cluster 2* Nilai Akademik

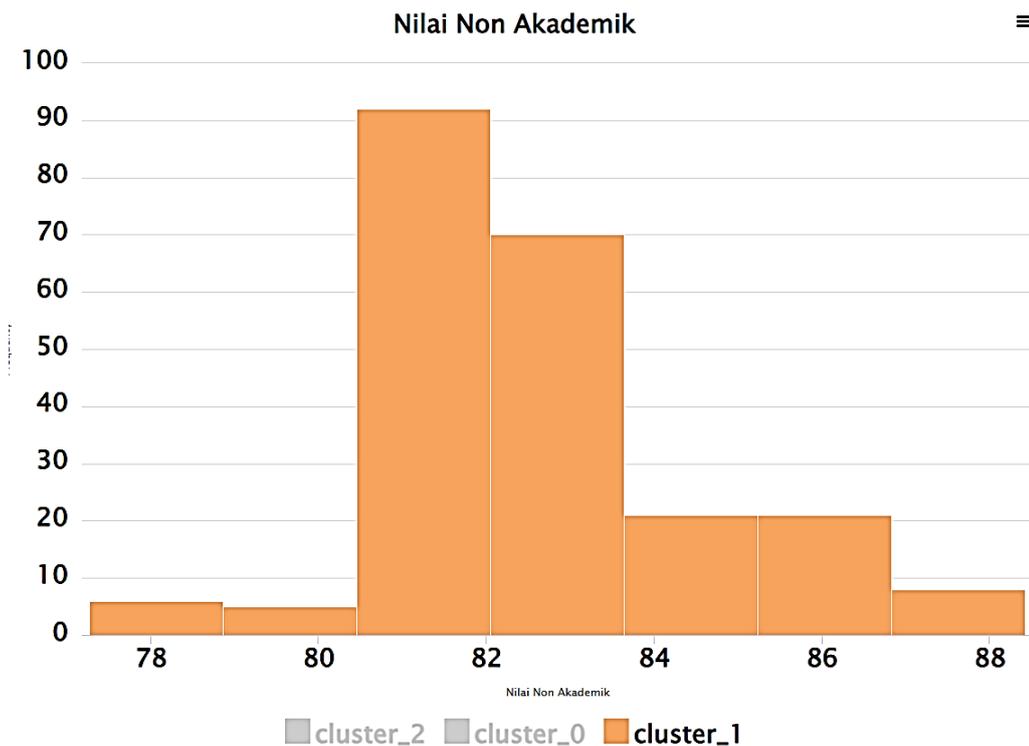
e. Nilai Non Akademik

Berdasarkan gambar visualisasi *cluster 0*, "Distribusi nilai non-akademik siswa dalam diagram menunjukkan bahwa terdapat 1 siswa dengan nilai non-akademik dalam rentang 75.8-77.394. Sebanyak 137 siswa memiliki nilai non-akademik dalam rentang 77.394-78.988. Jumlah siswa yang signifikan, yaitu 263 siswa, berada dalam rentang nilai 78.988-80.582. Selain itu, terdapat 59 siswa dengan nilai non-akademik dalam rentang 80.582-82.176, dan hanya 1 siswa yang memiliki nilai non-akademik dalam rentang 82.176-85.364. Data ini menunjukkan bahwa sebagian besar siswa memiliki nilai non-akademik antara 77.394 hingga 80.582, dengan puncak jumlah siswa berada pada nilai 78.988 hingga 80.582. Informasi lebih rinci mengenai distribusi nilai dalam *cluster* ini dapat dilihat pada Gambar 4.37



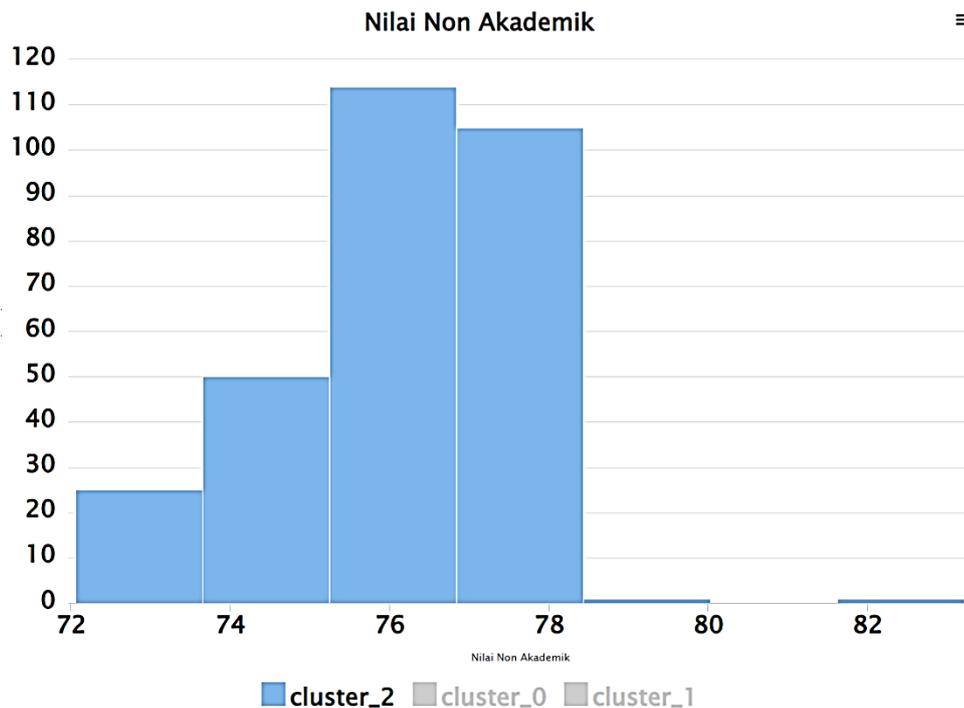
Gambar 4. 37 Visualisasi *Cluster 0* Nilai Non Akademik

Berdasarkan gambar Visualisasi *Cluster 1*, distribusi nilai non-akademik siswa dalam diagram menunjukkan bahwa terdapat 6 siswa dengan nilai non-akademik dalam rentang 77.27-78.864, 5 siswa dengan nilai non-akademik dalam rentang 78.864-80.458, dan 137 siswa memiliki nilai non-akademik dalam rentang 77.394-78.988. Jumlah siswa yang signifikan, yaitu 263 siswa, berada dalam rentang nilai 78.988-80.582. Selain itu, terdapat 92 siswa dalam rentang nilai 80.458-82.052, 59 siswa dalam rentang 80.582-82.176, dan 70 siswa dalam rentang 82.052-83.646. Terdapat juga 21 siswa dengan nilai non-akademik dalam rentang 83.646-85.24, 21 siswa dalam rentang 85.24-86.834, dan 8 siswa dalam rentang 86.834-88.428. Hanya 1 siswa yang memiliki nilai non-akademik dalam rentang 82.176-85.364. Data ini menunjukkan bahwa sebagian besar siswa memiliki nilai non-akademik antara 77.27 hingga 83.646, dengan puncak jumlah siswa berada pada nilai 78.988 hingga 80.582. Visualisasi data ini dapat dilihat pada Gambar 4.38



Gambar 4. 38 Visualisasi *Cluster 1* Nilai Non Akademik

Berdasarkan gambar Visualisasi *Cluster 2*, distribusi nilai non-akademik siswa dalam diagram menunjukkan bahwa terdapat 25 siswa dengan nilai non-akademik dalam rentang 72.06-73.654, 50 siswa dalam rentang 73.654-75.248, dan 114 siswa memiliki nilai non-akademik dalam rentang 75.248-76.842. Selain itu, terdapat 105 siswa dengan nilai non-akademik dalam rentang 76.842-78.436. Hanya 1 siswa yang memiliki nilai non-akademik dalam rentang 78.436-80.03, dan 1 siswa lagi dalam rentang 81.624-83.218. Data ini menunjukkan bahwa sebagian besar siswa memiliki nilai non-akademik antara 72.06 hingga 78.436, dengan puncak jumlah siswa berada pada nilai 75.248 hingga 76.842. Visualisasi data ini dapat dilihat secara lebih jelas pada Gambar 4.39



Gambar 4. 39 Visualisasi *Cluster 2* Nilai Non Akademik

4.7 Hasil Penelitian (Revisi dan Penambahan)

Berdasarkan hasil pengelompokan menggunakan metode *K-Means Clustering* menghasilkan tiga *cluster* utama yang menggambarkan kelompok siswa berdasarkan atribut jenis kelamin, jurusan, nilai akademik, nilai non-akademik, dan status alumni. Masing-masing *cluster* memberikan *insight* penting sebagai dasar perumusan strategi peningkatan mutu sekolah.

A. Analisis Visualisasi dan *Interpretasi Cluster*

1. Visualisasi Berdasarkan Jenis Kelamin

- *Cluster 0* dan *Cluster 2* didominasi oleh siswa perempuan, sedangkan *Cluster 1* memiliki distribusi yang lebih seimbang.
- *Interpretasi:* Ini menunjukkan bahwa siswa perempuan lebih tersebar pada kelompok performa sedang dan rendah. Sekolah dapat

mempertimbangkan program *mentoring* berbasis gender untuk mengoptimalkan potensi akademik dan non-akademik.

2. Visualisasi Berdasarkan Jurusan

- Seluruh cluster didominasi oleh jurusan IPA. Namun, pada **Cluster 2** jumlah siswa IPS sedikit lebih tinggi dibandingkan dua cluster lainnya.
- **Interpretasi:** Perlu ada strategi khusus dalam pendekatan pengajaran siswa IPS di **Cluster 2** yang cenderung memiliki performa rendah.

3. Visualisasi Berdasarkan Status Alumni

- **Cluster 1** paling banyak diisi oleh siswa yang melanjutkan kuliah, sedangkan **Cluster 0** dan **Cluster 2** didominasi oleh siswa yang langsung bekerja.
- **Interpretasi:** Siswa dengan performa tinggi lebih cenderung melanjutkan pendidikan. Sekolah perlu mengadakan program motivasi dan pelatihan kesiapan kuliah untuk siswa di cluster lain.

4. Visualisasi Berdasarkan Nilai Akademik

- **Cluster 1** memiliki nilai akademik tertinggi (rata-rata > 82),
- **Cluster 0** nilai sedang (rata-rata sekitar 79), dan
- **Cluster 2** berada pada nilai rendah (rata-rata < 78).
- **Interpretasi:** Sekolah dapat menerapkan program remedial untuk siswa **Cluster 2**, serta program akselerasi untuk siswa **Cluster 1**.

5. Visualisasi Berdasarkan Nilai Non-Akademik

- **Cluster 1** juga menunjukkan nilai non-akademik yang tinggi, menunjukkan siswa ini aktif dalam kegiatan di luar pelajaran.
- **Interpretasi:** Potensi siswa di Cluster 1 dapat dikembangkan melalui pelatihan kepemimpinan atau pengembangan karakter untuk menjadi duta sekolah.

B. Korelasi *Cluster*

1. Korelasi Nilai Akademik dengan Mutu Pendidikan

Cluster 1 memiliki hubungan positif dengan peningkatan mutu pendidikan, terlihat dari rata-rata nilai akademik tertinggi (82.892) dan kecenderungan lebih besar untuk melanjutkan studi ke perguruan tinggi (Status Alumni: 0.507). Sebaliknya, *Cluster 2* mencatat nilai akademik terendah (76.118), dengan mayoritas siswanya langsung memasuki dunia kerja setelah lulus (Status Alumni: 0.483).

Untuk meningkatkan mutu pendidikan, sekolah dapat menerapkan bimbingan akademik intensif bagi siswa di *Cluster 2* melalui kelas remedial dan pendampingan belajar, guna meningkatkan pencapaian akademik mereka. Sementara itu, bagi siswa *Cluster 1*, kelas akselerasi dapat menjadi strategi efektif untuk mengoptimalkan potensi mereka dan memberikan tantangan akademik yang lebih sesuai.

2. Korelasi Nilai Non Akademik dengan Pengembangan Karakter

Siswa *Cluster 1* menunjukkan nilai non-akademik tertinggi (82.716), yang mengindikasikan bahwa mereka aktif dalam kegiatan ekstrakurikuler dan memiliki *soft skills* yang baik. Siswa *Cluster 2*, dengan nilai non-akademik terendah (75.990), mungkin kurang terlibat dalam aktivitas pengembangan diri di luar akademik. Strategi peningkatan mutunya adalah Meningkatkan partisipasi ekstrakurikuler bagi siswa *Cluster 2* dengan pelatihan kepemimpinan dan sertifikasi keterampilan dan mengoptimalkan siswa *Cluster 1* sebagai mentor atau duta sekolah dalam kegiatan sosial dan akademik.

3. Korelasi Status Alumni dengan Kesuksesan Akademik

Siswa dalam *Cluster 1* menunjukkan kecenderungan terbesar untuk melanjutkan pendidikan ke jenjang perguruan tinggi (0.507), sementara sebagian besar siswa di *Cluster 0* dan *Cluster 2* memilih langsung memasuki dunia kerja setelah lulus.

Untuk meningkatkan mutu pendidikan, sekolah dapat menyelenggarakan bimbingan karier dan pelatihan persiapan SNBT bagi siswa di *Cluster* 0 dan 2 agar lebih siap menghadapi tantangan akademik maupun profesional. Selain itu, diperlukan program motivasi belajar serta penguatan *soft skill* untuk membantu siswa yang belum menentukan pilihan setelah kelulusan, sehingga mereka memiliki lebih banyak opsi dan kesiapan dalam mengambil keputusan terkait masa depan mereka.

4. Korelasi Jenis Kelamin dengan Performa Akademik dan Non Akademik

Siswa perempuan lebih dominan di *Cluster* 0 dan *Cluster* 2 (0.615 - 0.659), sedangkan *Cluster* 1 memiliki distribusi yang lebih seimbang. Pola ini menunjukkan bahwa siswa perempuan cenderung lebih banyak berada di kelompok dengan performa akademik dan non-akademik yang sedang atau rendah.

Untuk meningkatkan mutu pendidikan, sekolah dapat mengembangkan program mentoring berbasis gender guna membantu siswa perempuan mengoptimalkan potensi mereka dalam bidang akademik maupun non-akademik. Selain itu, perlu diterapkan metode pembelajaran yang lebih adaptif, yang mempertimbangkan pola performa berdasarkan gender, sehingga strategi belajar lebih sesuai dengan kebutuhan dan karakteristik siswa.

5. Korelasi Jurusan dengan Performa Akademik

Jurusan IPA mendominasi semua *cluster*, tetapi *Cluster* 2 memiliki sedikit lebih banyak siswa dari jurusan IPS dibandingkan *cluster* lainnya (0.385 vs 0.377). Hal ini menunjukkan bahwa siswa IPS mungkin menghadapi tantangan akademik yang lebih besar dibandingkan dengan siswa IPA.

Untuk meningkatkan mutu pendidikan, sekolah dapat menerapkan pendekatan pembelajaran yang lebih fleksibel bagi siswa IPS, seperti metode diskusi interaktif atau studi kasus yang relevan dengan bidang mereka. Selain itu, pelatihan keterampilan berbasis profesi dapat diberikan untuk meningkatkan kesiapan

mereka dalam memasuki dunia kerja, sehingga mereka memiliki kompetensi yang lebih kuat sesuai dengan kebutuhan industri.

C. Rangkuman Strategi Berdasarkan *Cluster*

- ***Cluster 0 (Performa Sedang)***: Fokus pada penguatan soft skill, pemberian motivasi belajar, serta pelatihan manajemen waktu.
- ***Cluster 1 (Performa Tinggi)***: Perlu disiapkan program lanjutan seperti pembinaan olimpiade, persiapan SNBT, dan pemberian beasiswa.
- ***Cluster 2 (Performa Rendah)***: Diperlukan intervensi belajar intensif, keterlibatan guru BK dan orang tua, serta pelatihan motivasi belajar.

Berdasarkan hasil analisis korelasi, SMA Negeri 1 Pagelaran dapat meningkatkan mutu pendidikan melalui strategi berbasis data, menyesuaikan program akademik dan non-akademik dengan karakteristik setiap *cluster*. Dengan penerapan intervensi yang tepat, sekolah dapat:

- Mengoptimalkan pembelajaran bagi siswa yang mengalami kesulitan akademik, memastikan mereka mendapatkan dukungan yang dibutuhkan.
- Memaksimalkan peran *ekstrakurikuler* dalam membentuk karakter dan meningkatkan kesiapan siswa menghadapi tantangan di masa depan.
- Memfasilitasi transisi yang lebih baik bagi siswa menuju dunia kerja atau pendidikan tinggi, sehingga mereka memiliki kesiapan yang lebih matang untuk melangkah ke jenjang berikutnya.

Pendekatan ini memungkinkan sekolah untuk lebih adaptif dalam mengembangkan program pendidikan yang sesuai dengan kebutuhan setiap siswa, menciptakan ekosistem belajar yang lebih efektif dan inklusif.