

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Pengumpulan Data

Data set pada penelitian ini diperoleh dari repository kagle mengenai kesehatan mental mahasiswa. Tujuan dari data set adalah untuk memberikan wawasan berharga tentang kesehatan mental mahasiswa dengan menangkap sejumlah faktor yang mungkin memengaruhi kesehatan mental mereka. Adapun data set pada penelitian ini terdiri dari 7022 data dan attribute berjumlah 20. Berikut attribute dalam data set kesehatan mental pada Mahasiswa yang ditunjukkan pada tabel 4.1.

**Tabel 4. 1** Atribute Data Set

Nama Atribute	Keterangan
Stress Level	Tingkat stres yang dialami oleh individu.
Depression Score	Skor yang merepresentasikan tingkat depresi yang dialami oleh individu.
Anxiety Score	Skor yang merepresentasikan tingkat kecemasan yang dialami oleh individu.
Sleep Quality	Kualitas tidur yang dialami oleh individu.
Physical Activity	Tingkat aktivitas fisik.
Diet Quality	Kualitas diet individu.
Social Support	Tingkat dukungan sosial yang diterima oleh individu.
Substance Use	Frekuensi penggunaan substansi seperti alkohol, rokok, atau obat-obatan lainnya.
Family_History	Apakah individu memiliki riwayat keluarga dengan masalah kesehatan mental.
Finansial stress	Tingkat stres keuangan yang dialami oleh individu (antara 0 dan 5).
Semester Credit Load	Jumlah skor kredit yang diambil oleh individu dalam semester (15-30).

Berikut visualisasi data set kesehatan mental mahasiswa pada rapidminer versi 10.1 seperti yang ditunjukkan pada gambar 4.1.

Row No.	Age	Course	Gender	CGPA	Stress_Level	Depression...	Anxiety_Sc...	Sleep_Qual...	Physical_A...
1	25	Others	Male	3.560	3	3	2	Good	Moderate
2	24	Engineering	Female	2.440	0	3	0	Average	Low
3	19	Business	Female	3.740	4	0	3	Good	Low
4	19	Computer Sci...	Male	3.850	2	1	0	Average	Low
5	18	Business	Male	3.400	3	3	4	Good	Low
6	21	Medical	Female	3.350	2	4	3	Good	Moderate
7	19	Law	Male	3.850	2	2	5	Good	Moderate
8	21	Business	Female	3.400	0	3	3	Average	Low
9	24	Medical	Male	3.800	3	2	1	Poor	Low
10	19	Engineering	Female	3.000	2	5	0	Average	Moderate
11	23	Law	Female	3.740	3	2	4	Average	Low

Gambar 4.1 Visualisasi data set pada rapidminer

## 4.2 Preprocessing Data

### 1. Cleaning data

Tahap cleaning data melibatkan algoritma replace missing value pada rapidminer. Algoritma ini memiliki fungsi berupa pembersihan data, penyesuaian data, normalisasi data, transformasi data, penanganan missing value, reduksi dimensi, maupun pemilihan subset fitur. Adapun pada saat input data set terdapat 27 missing data masing-masing 15 data attribute substance use dan 12 data pada attribute CGPA seperti yang ditampilkan pada gambar 4.2.

✓ Diet_Quality	Nominal	3	Poor (1369)	Average (4288)	Average (4288), Good (1365)
✓ Social_Support	Nominal	3	Low (1376)	Moderate (3470)	Moderate (3470), High (2175)
✓ Relationship_Status	Nominal	2	Married (1369)	Single (3574)	Single (3574), In a Relation...
✓ Substance_Use	Nominal	15	Frequently (405)	Never (5903)	Never (5903), Occasionally (...
✓ Counseling_Service_Use	Nominal	3	Frequently (575)	Never (4263)	Never (4263), Occasionally (...
✓ Course	Nominal	0	Others (715)	Medical (2105)	Medical (2105), Law (1385)
✓ Gender	Nominal	0	Female (3475)	Male (3047)	Male (3547), Female (3475)
✓ CGPA	Real	12	2.440	4	3.491
✓ Stress_Level	Integer	0	0	5	2.428
✓ Depression_Score	Integer	0	0	5	2.254

Gambar 4. 2 Data Missing Value

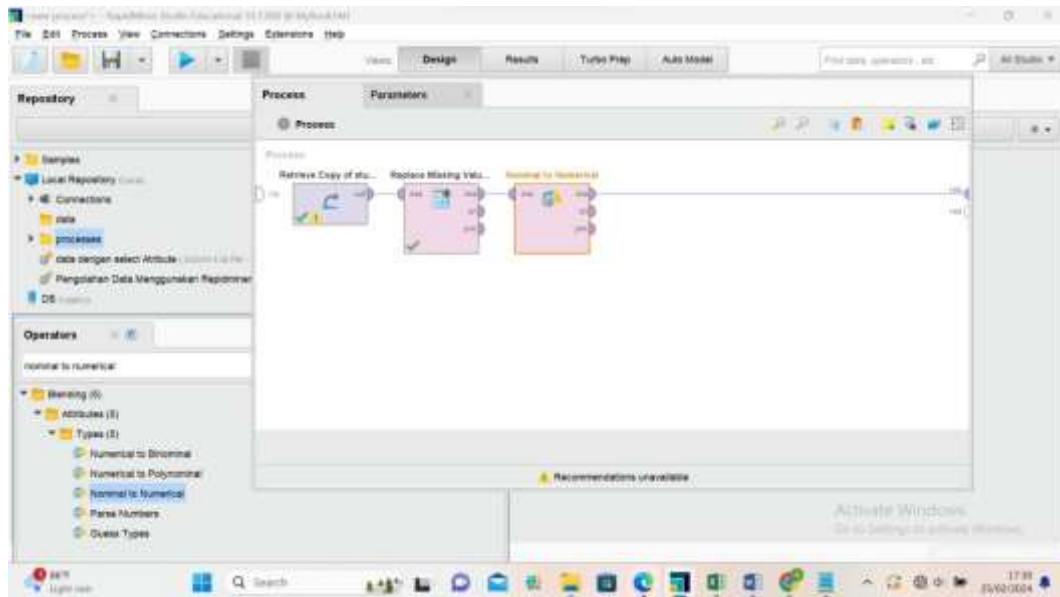
Pada tahap ini data disajikan secara statistik sebagai hasil tahapan pada cleaning data. Berdasarkan hasil pada cleaning data, semua attribute tidak ada data missing atau data kosong. Sehingga berpotensi untuk menghasilkan model klasterisasi yang maksimal. Berikut visualisasi data hasil tahapan cleaning data yang ditampilkan pada gambar 4.3.

Course	Polynomial	0	Others (715)	Medical (2105)	Medical (2105), Law (1385)
Gender	Polynomial	0	Female (3475)	Male (3547)	Male (3547), Female (3475)
CGPA	Real	0	2.440	4	3.491
Stress_Level	Integer	0	0	5	2.428
Depression_Score	Integer	0	0	5	2.254
Diet_Quality	Polynomial	0	Poor (1368)	Average (4268)	Average (4268), Good (1385)
Social_Support	Polynomial	0	Low (1378)	Moderate (3470)	Moderate (3470), High (2178)
Relationship_Status	Polynomial	0	Married (1368)	Single (3574)	Single (3574), In a Relationship (1368)
Substance_Use	Polynomial	0	Frequently (405)	Never (5918)	Never (5918), Occasionally (405)
Counseling_Service_Use	Polynomial	0	Frequently (575)	Never (4263)	Never (4263), Occasionally (575)

Gambar 4. 3 Hasil Cleaning Data

## 2. Tahap transformasi data

Pada tahap ini data diubah atau digabung ke dalam format yang sesuai dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Pada konteks data set kesehatan mental mahasiswa terdapat beberapa kategori data berupa nominal, numerik, kategorikal bahkan binomial. Untuk memaksimalkan pada tahapan performance atau perhitungan dalam menentukan kluster terbaik. Maka saya menggunakan algoritma *nominal to numerical* untuk membuat data menjadi homogen. Berikut penggunaan algoritma *nominal to numerical* pada rapidminer seperti yang ditunjukkan pada gambar 4.4.



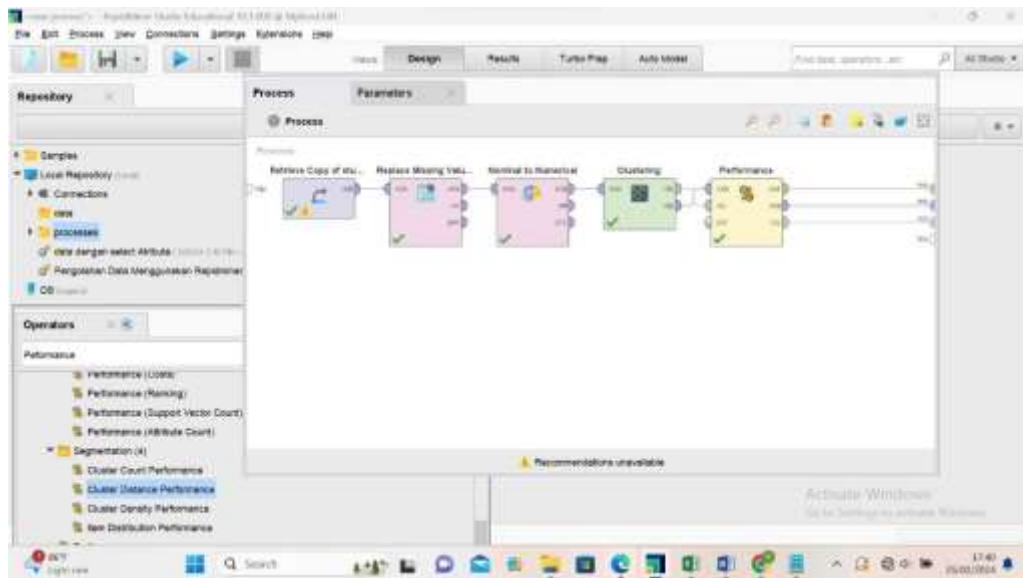
Gambar 4. 4 Transformasi data pada rapidminer

### 4.3 Tahapan K-Means Clustering

Teknik clustering merupakan suatu pendekatan analisis data yang bertujuan untuk mengelompokkan entitas data secara otomatis berdasarkan kemiripan karakteristiknya, tanpa adanya label kelas yang telah ditentukan sebelumnya. Clustering, seringkali diidentifikasi sebagai metode unsupervised learning, memberikan kemampuan untuk mengklasifikasikan data yang belum memiliki label kelas yang jelas atau terdefinisi. Metode ini menjadi esensial dalam konteks analisis data kompleks, di mana informasi tentang struktur kelas tidak diketahui atau sulit ditentukan. Proses clustering melibatkan identifikasi pola-pola intrinsik dalam data, dengan tujuan membentuk kelompok-kelompok homogen yang memiliki tingkat kemiripan tinggi di antara anggotanya.

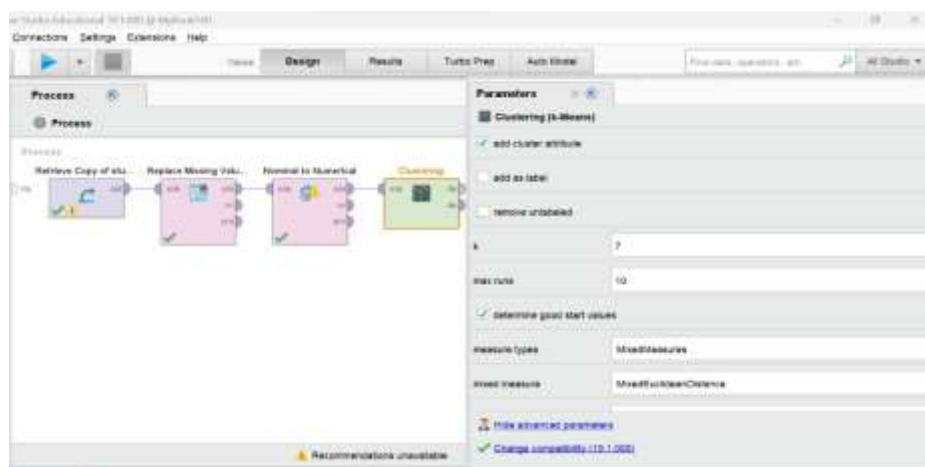
Adapun pada penelitian ini, saya menggunakan klasterisasi metode k-means. Algoritma ini bekerja dengan menjalankan perintah banyak kelas klaster ( $k$ ) yang diinginkan. Selain itu, k-means bekerja dengan menunggu jumlah pengulangan eksekusi perintah, yang dikenal dengan istilah iterasi, maksimum dan sebuah nilai konstanta yang digunakan untuk penentuan kapan iterasi dihentikan. Berikut

penggunaan algoritma k-means pada penelitian ini seperti yang ditampilkan pada gambar 4.5.



**Gambar 4. 5** Tahapan K-Means Clustering

Pada parameter algoritma clustering, saya melakukan simulasi sebanyak 7 kali pembentukan kelas untuk mengetahui model terbaik. Adapun max run pada parameter saya sesuaikan dengan default 10 kali untuk menentukan konsistensi dari model. Pada measures type, saya menggunakan mixed measures karena type data pada data set yang heterogen. Berikut pengaturan pada parameter algoritma k-means perhitungan data set kesehatan mental mahasiswa seperti yang ditunjukkan pada gambar 4.6.



**Gambar 4. 6** Pengaturan Parameter Aloritma K-Means

Adapun hasil dari pembentukan kluster dari nilai  $k = 2$  sampai  $k = 7$  sebagai berikut:

k-2	Hasil dari model klustering pada RapidMiner menghasilkan dua kluster utama. Kluster 0, dengan jumlah 3268 item, dan Kluster 1, dengan jumlah 3754 item, memberikan gambaran mengenai pemisahan data menjadi dua kelompok yang signifikan. Kluster ini mencerminkan pola-pola atau karakteristik tertentu yang membedakan antara dua kelompok tersebut. Kluster dengan jumlah item terendah adalah Kluster 0, sedangkan Kluster 1 memiliki jumlah item tertinggi. Perbedaan jumlah item ini mengindikasikan bahwa Kluster 1 mungkin memiliki variasi atau kompleksitas yang lebih tinggi dibandingkan dengan Kluster 0
k-3	Cluster 1 dengan 1570 item, dan Cluster 2 dengan 2818 item. Analisis lebih lanjut mengungkapkan bahwa Cluster 2 merupakan kluster tertinggi dengan jumlah item terbanyak, sedangkan Cluster 1 adalah kluster dengan jumlah item terendah
k-4	Kluster 1 menjadi kluster dengan jumlah data terbanyak, mencapai 2283 item, mengindikasikan sekelompok besar data yang memiliki kesamaan karakteristik tertentu. Kluster dengan jumlah data terendah adalah Kluster 0, yang terdiri dari 1079 item. Setiap kluster mewakili kelompok data yang memiliki tingkat kemiripan yang tinggi di antara anggotanya, sementara perbedaan karakteristik antar-kluster dapat diidentifikasi.
k-5	Kluster 0 membawahi sebanyak 1644 entitas data, Kluster 1 terdiri dari 1016 entitas data, sementara Kluster 2 mencakup 1635 entitas data. Kluster 3 memiliki 1158 entitas data, dan Kluster 4 terdiri dari 1569 entitas data.
k-6	Cluster 0 memiliki 1190 item, Cluster 1 mencakup 1397 item, sementara Cluster 2 terdiri dari 1535 item. Cluster 3 memiliki 1167 item, Cluster 4 terdiri dari 510 item, dan Cluster 5 mencakup 1223 item

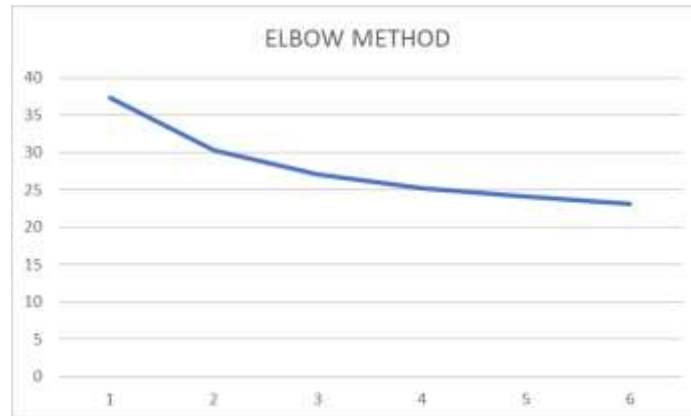
k-7	Cluster 0 terdiri dari 724 item, sementara Cluster 1 memiliki 1135 item. Selanjutnya, Cluster 2 mencakup 1223 item, Cluster 3 terdiri dari 481 item, dan Cluster 4 memiliki 871 item. Cluster 5 menampilkan jumlah item tertinggi dengan 1534, sedangkan Cluster 6 dan Cluster 7 masing-masing memiliki 1054 item
-----	---

#### 4.4 Tahap Analisis Kluster

Analisis kluster merupakan salah satu teknik pada data mining yang digunakan untuk menemukan kluster-kluster dari himpunan data secara otomatis atau semi-otomatis. Berdasarkan pendekatan dan/atau konsep yang diadopsi, teknik clustering dapat dikategorikan ke dalam metoda yang berbasis partisi, hirarki, dan densitas/kerapatan, grid, model dan konstrain. Adapun pada penelitian ini saya menggunakan simulasi model untuk menentukan kluster terbaik pada data set kesehatan mental mahasiswa. Berikut hasil simulasi penentuan kluster:

Banyak kluster (k)	Hasil
2	37.318
3	30.29
4	27.04
5	25.234
6	24.084
7	23.168

Untuk menentukan nilai kluster (k) terbaik pada data set Kesehatan mental mahasiswa. Saya menggunakan metode elbow. Metode siku (Elbow Method) bertujuan untuk menemukan jumlah kluster yang optimal sehingga hasil klustering menjadi sesuai dengan struktur intrinsik dari data tanpa overfitting atau underfitting. Berikut penggunaan metode elbow (siku) seperti yang ditunjukkan pada gambar 4.7.



**Gambar 4. 7** Metode Elbow

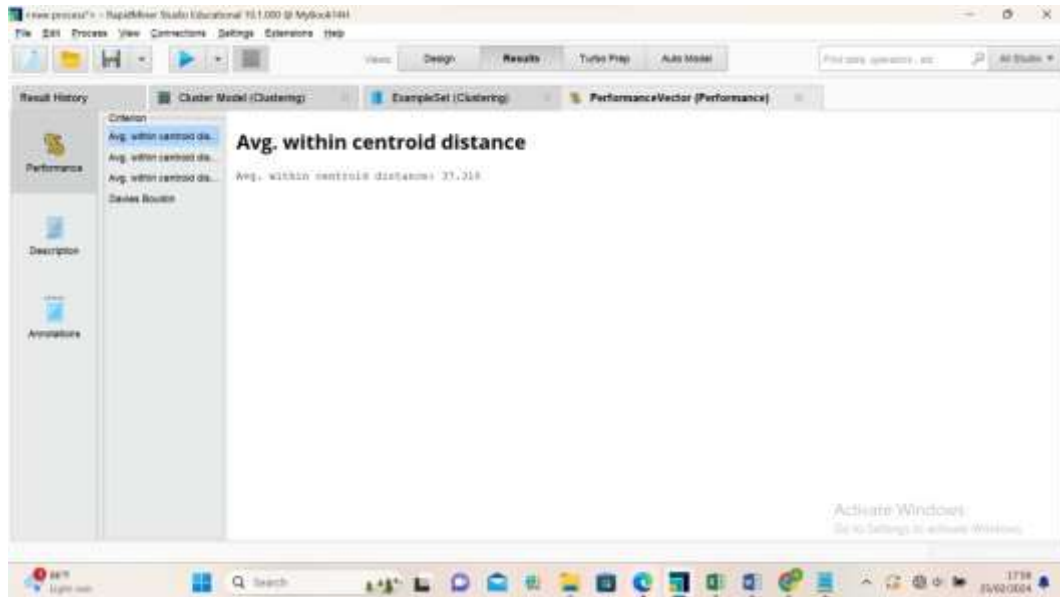
Berdasarkan visualisasi pada elbow method di atas, maka diketahui bahwa kelengkungan tertajam sehingga membentuk sudut siku berada pada kluster 2. Sehingga ditentukan bahwa analisis kluster terbaik berada pada kluster atau nilai  $k = 2$

#### 4.5 Hasil Clustering

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

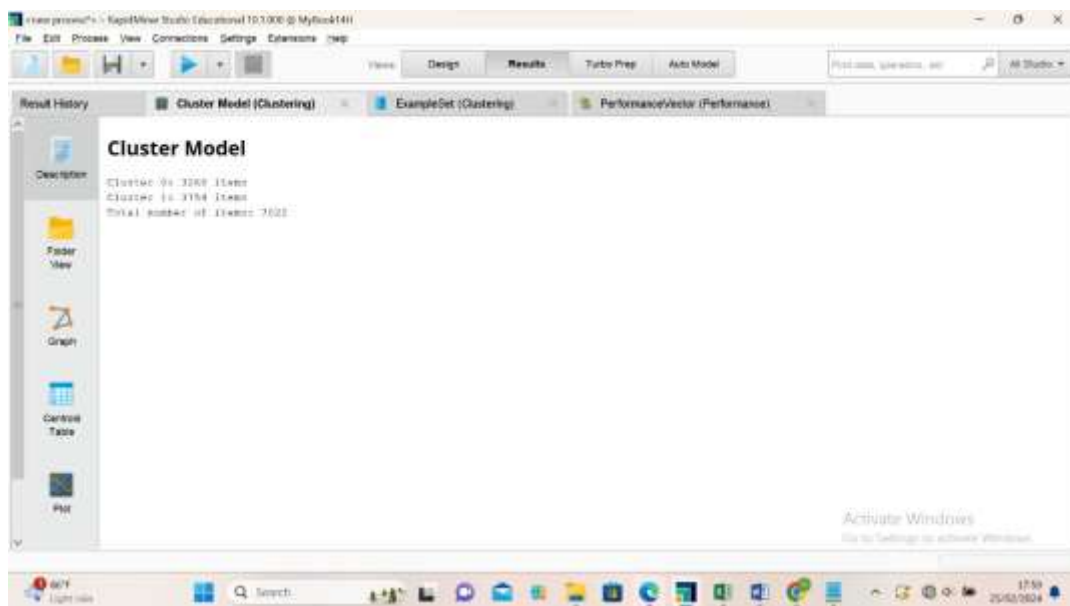
1. Berdasarkan Analisa data set kesehatan mental mahasiswa memiliki nilai kluster terbaik  $k = 2$ . Adapun nilai Avg. within centroid distance adalah 37.318. nilai Avg. within centroid distance: 37.318, dapat diartikan bahwa rata-rata jarak antara titik data dalam kluster dengan pusat kluster adalah sekitar 37.318. Semakin kecil nilai ini, semakin padat dan homogen kluster tersebut, berikut adalah performance pada tampilan repidminer yang ditunjukkan pada gambar 4.8.





**Gambar 4. 8** hasil performance terbaik

Adapun rincian masing-masing kluster dibagi menjadi Kluster 0 dengan jumlah 3268 item dan Kluster 1 dengan jumlah 3754 item. Berikut gambar cluster model yang terbentuk seperti gambar 4.8.



**Gambar 4. 9** Hasil Cluster Model

Berdasarkan hasil klusterisasi, maka data set kesehatan mental mahasiswa dengan jumlah 7022 dibagi menjadi 2 karakteristik kemiripan data. Sebanyak 3268 memiliki karakteristik data yang sama dan termasuk dalam kluster 0. Sedangkan 3754 data memiliki karakteristik yang mirip serta termasuk dalam kluster 1.

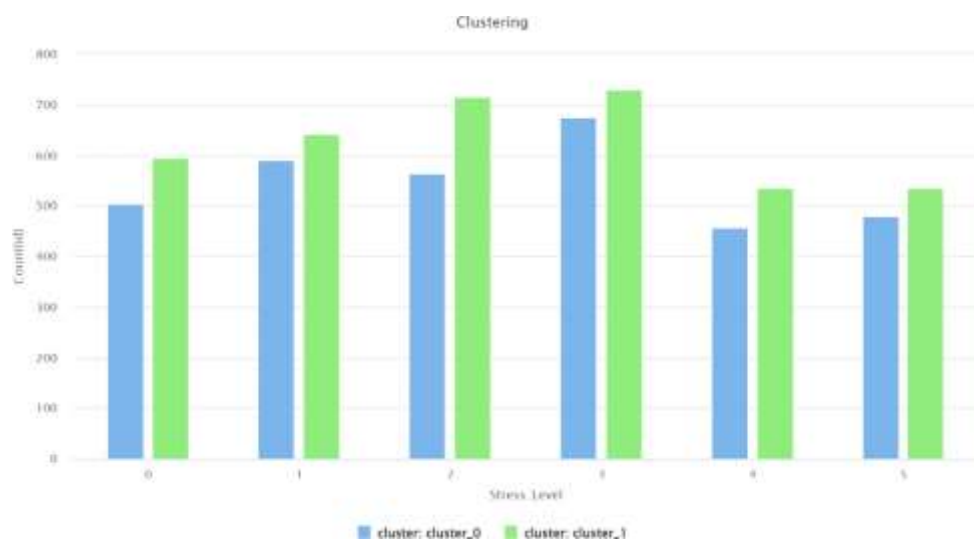
Untuk menentukan karakteristik kluster 0 dan kluster 1 dalam menentukan penggolongan kondisi kesehatan mental mahasiswa. Maka saya menggunakan tiga indikator utama berupa stress level, anxiety, dan depression sebagai berikut:

### 1. Stress level

Adapun hasil klasterisasi pada stress level digambarkan pada table berikut:

Stress Level	Cluster 0	Cluster 1
0 (Tidak ada Stress)	504	594
1 (Sangat Rendah)	590	643
2 (Sedikit Stres)	564	715
3 (Stress sedang)	674	731
4 (Stres tinggi)	457	535
5 (Stres sangat tinggi)	479	536

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih rendah dari pada cluster 1 pada indikator stress level. Pada setiap golongan pada stress level, cluster 0 lebih kecil dari pada cluster 1. Berikut visualisasi dari cluster 0 dan cluster 1 untuk Tingkat kecemasan mahasiswa yang ditunjukkan pada gambar 4.10.



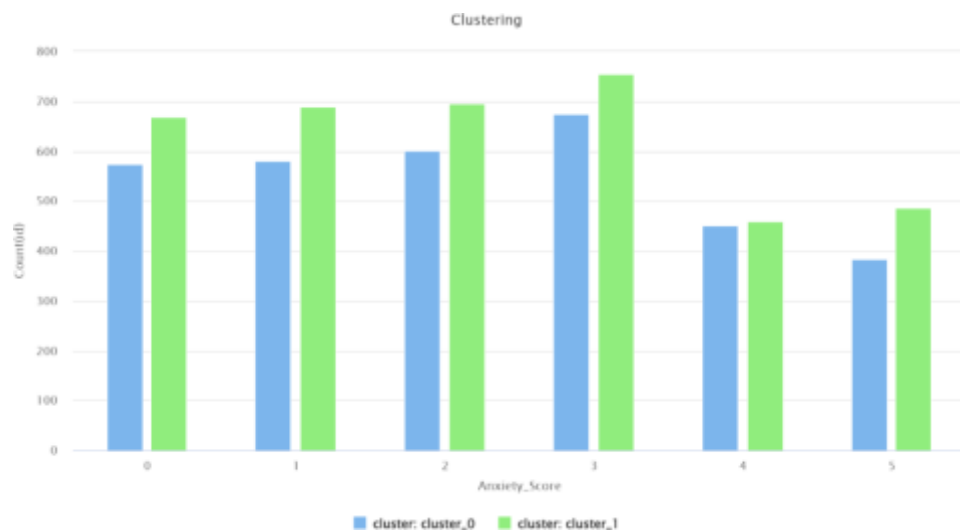
**Gambar 4.10** Visualisasi Stress Level

## 2. Anxiety Score

Berikut hasil klasterisasi pada anxiety score atau tingkat kecemasan:

Anxiety Score	Cluster 0	Cluster 1
0 (Tidak ada kecemasan)	504	594
1 (Sangat Rendah)	590	643
2 (Sedikit Kecemasan)	564	715
3 (Kecemasan sedang)	674	731
4 (Kecemasan tinggi)	457	535
5 (Kecemasansangat tinggi)	479	536

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih rendah dari pada cluster 1 pada indikator kecemasan. Pada setiap golongan pada anxiety level, cluster 0 lebih kecil dari pada cluster 1. Berikut visualisasi dari cluster 0 dan cluster 1 untuk Tingkat kecemasan mahasiswa yang ditunjukkan pada gambar 4.11.



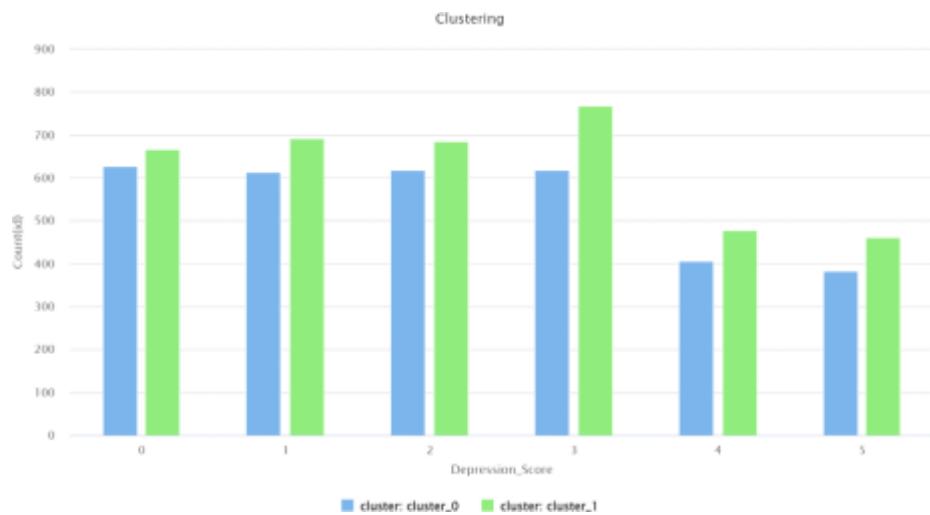
**Gambar 4. 11** Visualisasi Anxiety Score

### 3. Depression

Berikut hasil klasterisasi pada Depression score atau tingkat depresi mahasiswa:

Depression Score	Cluster 0	Cluster 1
0 (Tidak ada depresi)	504	594
1 (Sangat Rendah)	590	643
2 (Sedikit Depresi)	564	715
3 (Depresi sedang)	674	731
4 (Depresi tinggi)	457	535
5 (Depresi sangat tinggi)	479	536

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih rendah dari pada cluster 1 pada indikator depresi. Pada setiap golongan padat ingkat depresi, cluster 0 lebih kecil dari pada cluster 1. Berikut visualisasi dari cluster 0 dan cluster 1 untuk Tingkat kecemasan mahasiswa yang ditunjukkan pada gambar 4.12.



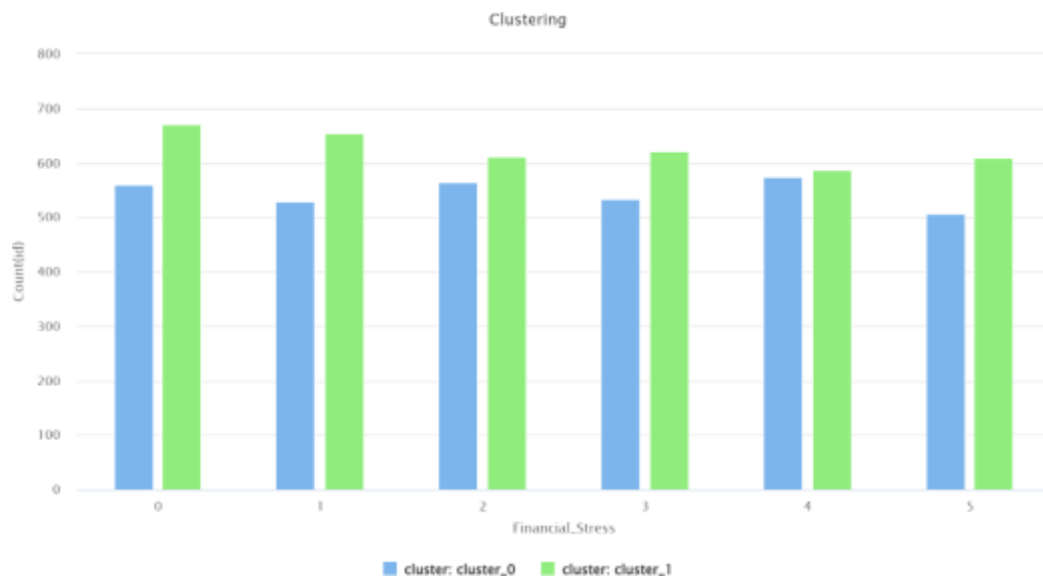
**Gambar 4.12** . Visualisasi Depression\_Score

#### 4. Financial\_Stress

Berikut adalah Berikut hasil klasterisasi pada financial stress.

Depresion Score	Cluster 0	Cluster 1
0 (Tidak ada stress keuangan)	504	594
1 (Sangat Rendah)	590	643
2 (Sedikit stress keuangan)	564	715
3 (Stress keuangan sedang)	674	731
4 (Stresess keuangan tinggi)	457	535
5 (Stress keuangan sangat tinggi)	479	536

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih rendah daripada cluster 1. Pada setiap golongan pada financial stress, cluster 0 lebih kecil dari pada cluster 1. Berikut visualisasi dari cluster 0 dan cluster 1 untuk financial stress yang ditunjukkan pada gambar 4.13.



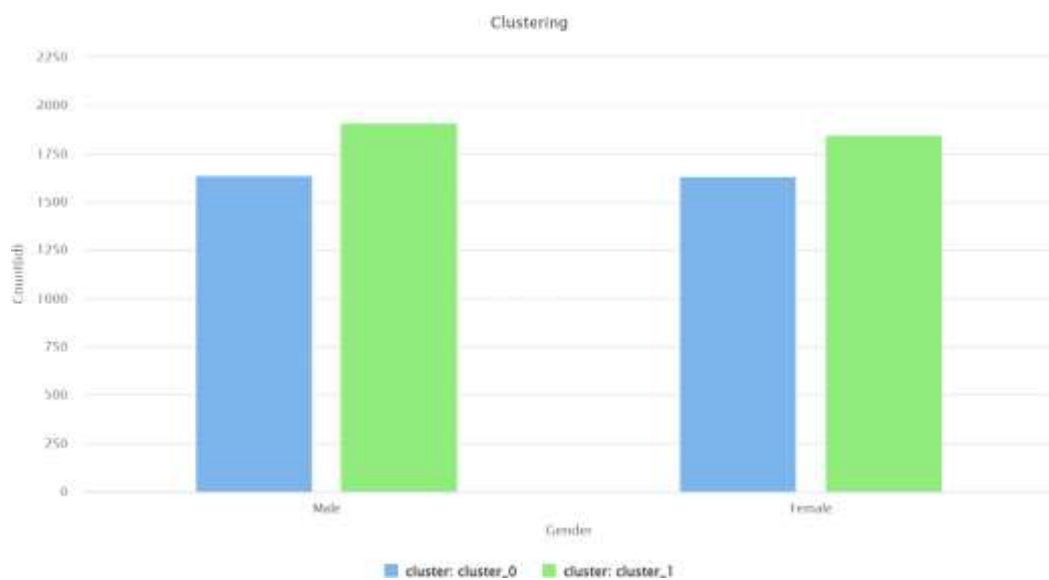
Gambar 4. 13 Gambar Visualisasi Financial\_Stress

## 5. Gender

Berikut adalah Berikut hasil klasterisasi pada Gender

Gender	Cluster 0	Cluster 1
Male	1638	1909
Female	1630	1845

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 maupun cluster 1 pada gender didominasi oleh laki laki. Berikut visualisasi dari cluster 0 dan cluster 1 untuk gender yang ditunjukkan pada gambar 4.14.



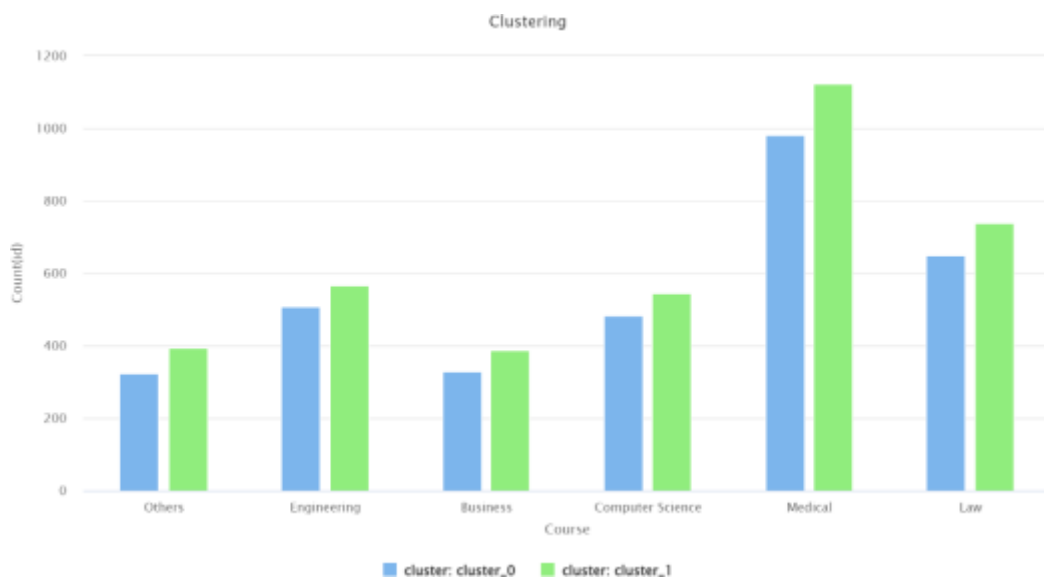
**Gambar 4. 14** Visualisasi Gender

## 6. Course

Berikut adalah hasil klasterisasi pada Course

Course	Cluster 0	Cluster 1
Other	322	393
Engineering	507	507
Bussiness	328	389
Computer Science	482	546
Medical	981	1124
Law	648	737

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 maupun cluster 1 pada Course didominasi oleh Medical. Berikut visualisasi dari cluster 0 dan cluster 1 untuk Course yang ditunjukkan pada gambar ditunjukkan pada gambar 4.15.



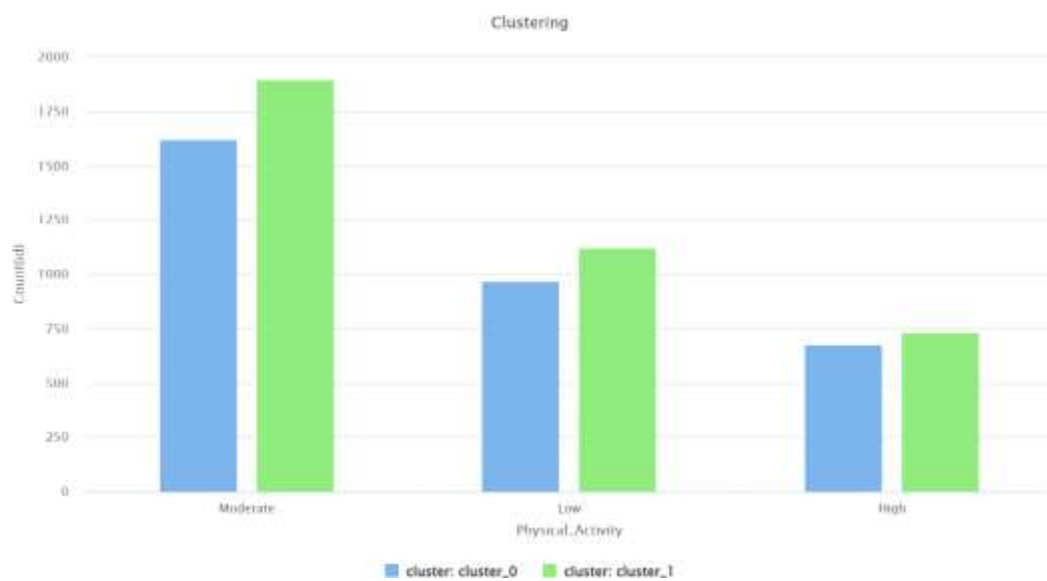
Gambar 4. 15 Visualisasi Course

## 7. Physical Activity

Berikut adalah hasil klasterisasi pada Physical Activity:

Physical Activity	Cluster 0	Cluster 1
Moderat	1622	1899
Low	969	1122
Hight	677	733

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih baik dari pada cluster 1 pada physical Activity . Pada setiap golongan pada physical activity, cluster 0 maupun cluster 1 memiliki nilai tertinggi adalah Moderat. Berikut visualisasi dari cluster 0 dan cluster 1 untuk physical Activity ditunjukkan pada gambar 4.16.



**Gambar 4.16** Visualisasi Physical Activity

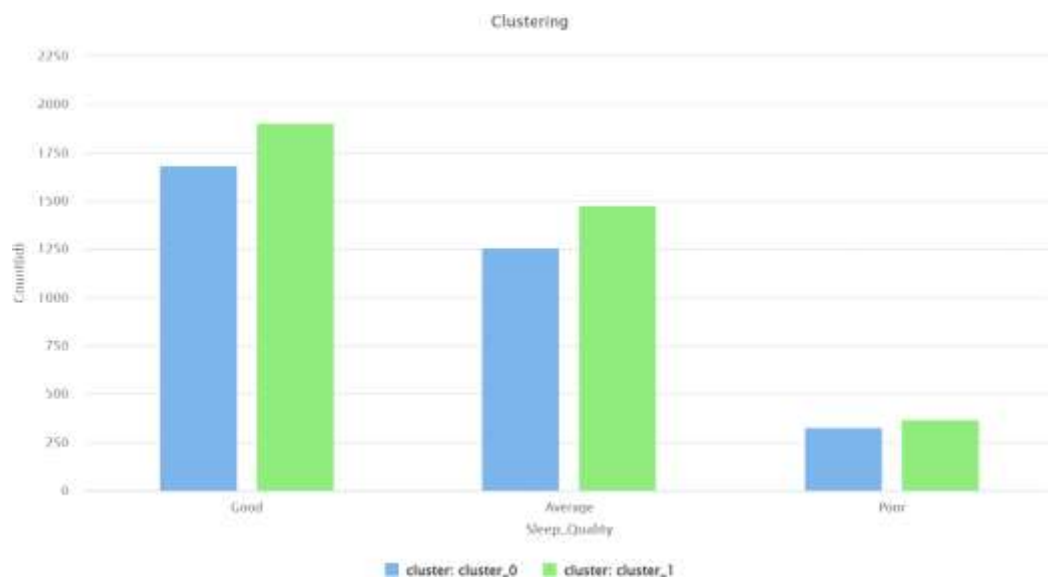


## 8. Sleep Quality

Berikut adalah hasil klasterisasi pada sleep quality

Sleep_Quality	Cluster 0	Cluster 1
Good	1683	1909
Average	1259	1476
Poor	326	372

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih baik dari pada cluster 1 pada indicator sleep quality. Pada setiap golongan pada physical activity, cluster 0 maupun cluster 1 memiliki nilai tertinggi adalah Good. Berikut visualisasi dari cluster 0 dan cluster 1 untuk sleep quality ditunjukkan pada gambar 4.17.



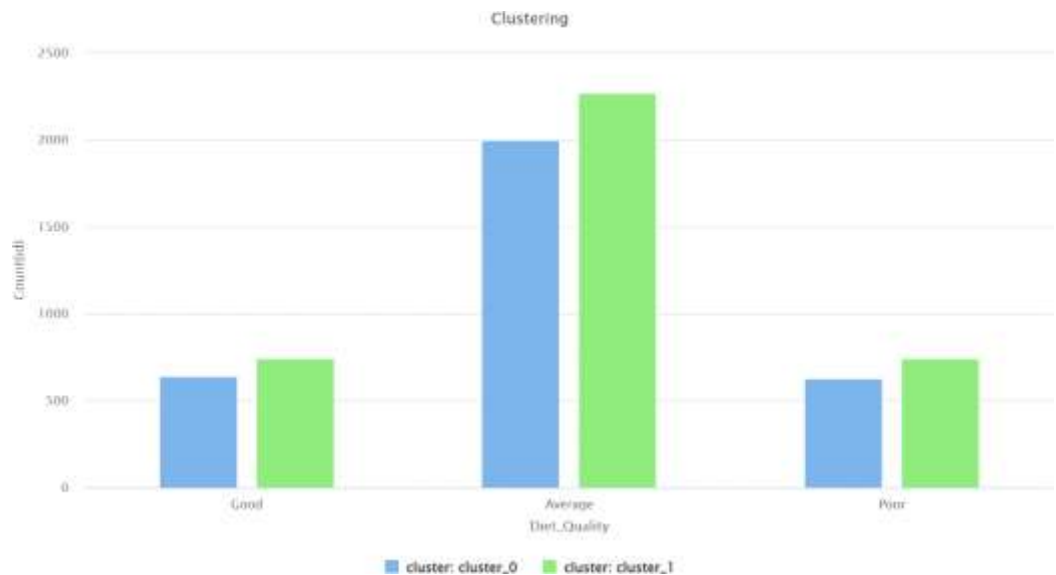
Gambar 4. 17 Visualisasi Sleep Quality

## 9. Diet Quality

Berikut adalah hasil klasterisasi pada Diet Quality:

Diet Quality	Cluster 0	Cluster 1
Good	641	744
Average	1999	2269
Poor	628	741

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih baik dari pada cluster 1 pada indicator diet quality. Pada setiap golongan pada physical activity, cluster 0 maupun cluster 1 memiliki nilai tertinggi adalah Average. Berikut visualisasi dari cluster 0 dan cluster 1 untuk sleep quality ditunjukkan pada gambar 4.18.



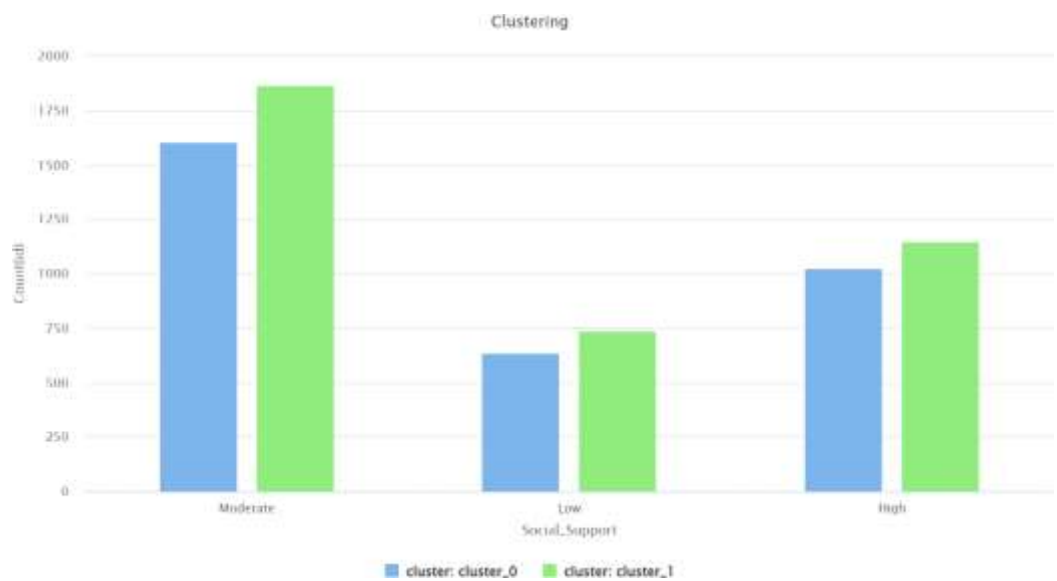
Gambar 4.18 Visualisasi Diet Quality

## 10. Sosial\_Support

Berikut adalah Berikut hasil klasterisasi sosial support:

Sosial_Support	Cluster 0	Cluster 1
Moderat	1603	1867
Low	637	739
High	1028	1148

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih baik dari pada cluster 1 pada indicator sosial support. Pada setiap golongan pada physical activity, cluster 0 maupun cluster 1 memiliki nilai tertinggi adalah Moderat. Berikut visualisasi dari cluster 0 dan cluster 1 untuk physical Activity ditunjukkan pada gambar 4.19.



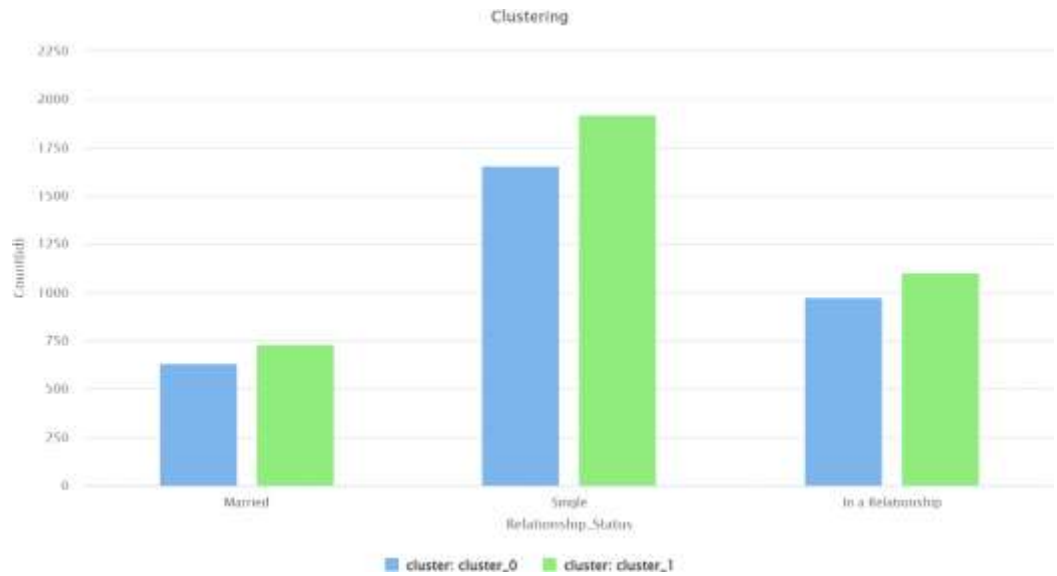
**Gambar 4.19** Visualisasi Social\_Support

## 11. Relationship\_Status

Berikut adalah Berikut hasil klasterisasi pada relationship status:

Relationship_Status	Cluster 0	Cluster 1
Married	636	733
Single	1655	1919
In a Relationship	977	1102

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui Pada setiap golongan pada Relationship status, cluster 0 maupun cluster 1 memiliki nilai tertinggi adalah Single. Berikut visualisasi dari cluster 0 dan cluster 1 untuk physical Activity ditunjukkan pada gambar 4.20.



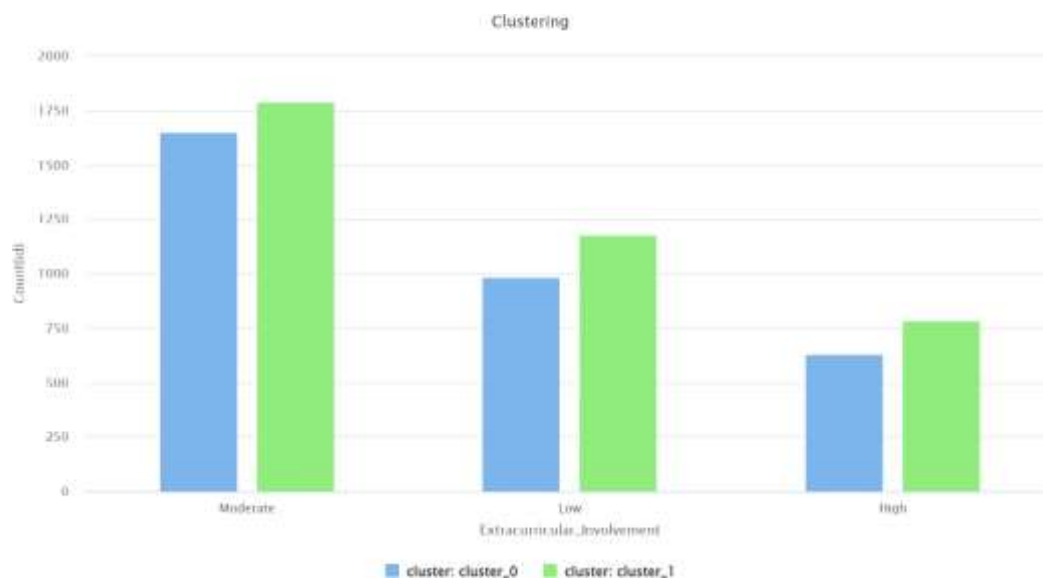
**Gambar 4.20** Visulisasi Relationship Status

## 12. Exctracurricular\_Involemen

Berikut adalah hasil klasterisasi pada Extracurricular Involemen:

Exctracurricular_Involemen	Cluster 0	Cluster 1
Moderat	1652	1788
Low	985	1179
Hight	631	787

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih baik dari pada cluster 1 pada Indicator Exctracurricular\_Involemen. Pada setiap golongan pada Exctracurricular, cluster 0 maupun cluster 1 memiliki nilai tertinggi adalah Moderat. Berikut visualisasi dari cluster 0 dan cluster 1 untuk Exctracurricular\_Involemen ditunjukkan pada gambar 4.19.



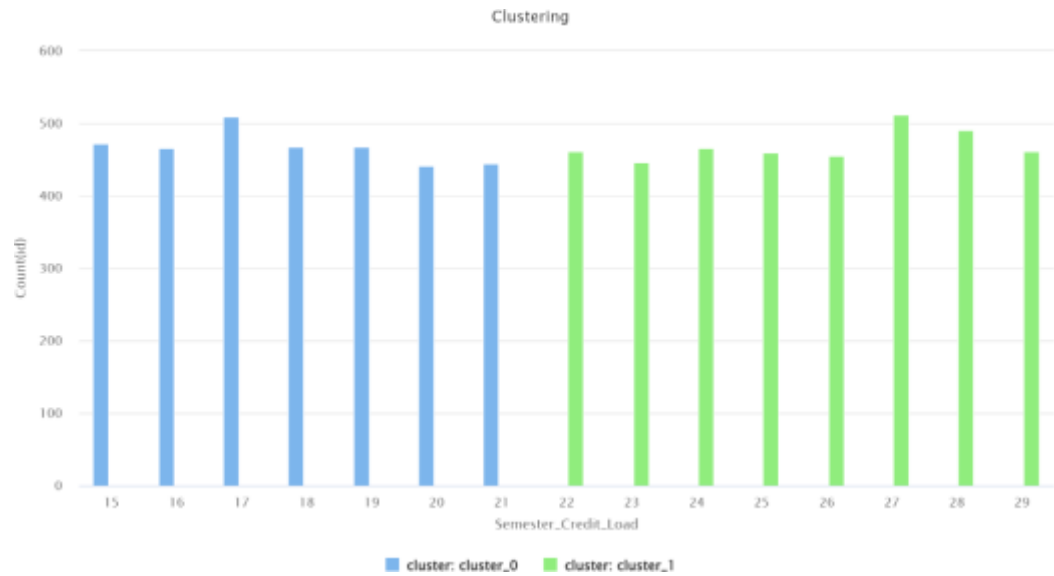
**Gambar 4.21** Visualisasi Extracurricular\_Involemen

### 13. Semester Credit Load

Berikut adalah hasil klasterisasi pada Semester Credit Load

Semester Credit Load	Cluster 0	Cluster 1
15-21	3268	0
22-29	0	3754

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa pada cluster 0 semester credit load yang diambil adalah 15-21 lalu untuk cluster 1 adalah 22-29. Berikut visualisasi dari cluster 0 dan cluster 1 untuk Semester Credit Load ditunjukkan pada gambar 4.20.



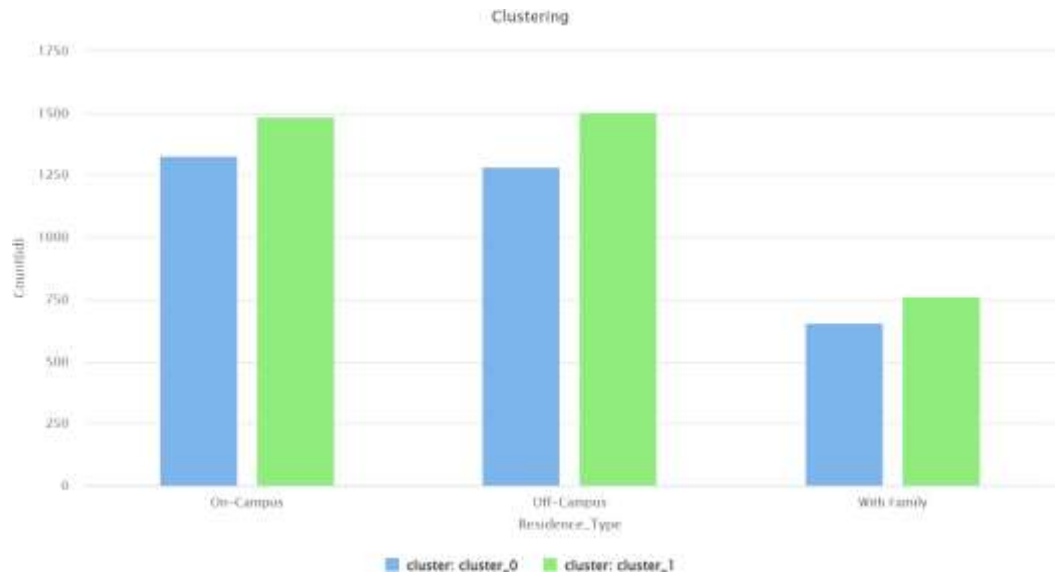
Gambar 4.20. Visualisasi Semester Credit Load

#### 14. Residen\_Type

Berikut adalah Berikut hasil klasterisasi pada Residen\_Type

Residen Type	Cluster 0	Cluster 1
On Campus	1328	1487
Off Campus	1285	1508
With Family	655	764

Berdasarkan perbandingan statistik pada populasi mahasiswa, diketahui bahwa cluster 0 lebih baik dari pada cluster 1 pada indicator residen type. Pada cluster 0 nilai tertinggi adalah On campus dan untuk cluster 1 nilai tertinggi adalah Off Campus. Berikut visualisasi Residen\_Type yang ditunjukkan pada gambar 4.21.



Gambar 4. 22 Visulisasi Residence Type