

ANALISIS CLUSTER MAHASISWA PENDIDIKAN MATEMATIKA UNIVERSITAS NIAS

by Lawuna Memori Putra

Submission date: 30-Oct-2023 12:51AM (UTC-0400)

Submission ID: 2211556731

File name: Memori_Putra_Lawuna.docx (230.39K)

Word count: 7714

Character count: 46183

**ANALISIS *CLUSTER* MAHASISWA PENDIDIKAN
MATEMATIKA UNIVERSITAS NIAS**

SKRIPSI



**Oleh
MEMORI PUTRA LAWUNA
NIM 192117040**

**6
PROGRAM STUDI PENDIDIKAN MATEMATIKA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS NIAS
2023**

**ANALISIS *CLUSTER* MAHASISWA PENDIDIKAN
MATEMATIKA UNIVERSITAS NIAS**

SKRIPSI

**Diajukan Kepada
Universitas Nias
Untuk Memenuhi Salah Satu Per syaratan Menyelesaikan
Program Sarjana Pendidikan**

**Oleh
MEMORI PUTRA LAWUNA
NIM 192117040**

**6
PROGRAM STUDI PENDIDIKAN MATEMATIKA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS NIAS
2023**

BAB I PENDAHULUAN

1.1 Latar Belakang

Pulau Nias saat ini memiliki beberapa Perguruan Tinggi dimana salah satunya adalah Universitas Nias. Universitas Nias adalah satu dari beberapa perguruan tinggi yang berada di Pulau Nias yang terdiri dari 3 Fakultas yakni Fakultas Keguruan dan Ilmu Pendidikan, Fakultas Ekonomi dan Fakultas Saintek. Universitas Nias didirikan pada tanggal 22 September 2021 berdasarkan SK Kemdikbud Ristekdikti Nomor 400/E/O/2021. Universitas Nias merupakan penggabungan Perguruan Tinggi IKIP Gunungsitoli menjadi Fakultas Keguruan dan Ilmu Pendidikan, STIE Pembnas menjadi Fakultas Ekonomi serta penambahan fakultas baru yakni Fakultas Sains dan Teknologi.

Pada Fakultas Keguruan dan Ilmu Pendidikan terdapat salah satu program studi yakni program studi pendidikan matematika dimana yang saat ini juga calon peneliti berada. Dalam Program Studi ini ada banyak data-data yang dapat diolah untuk dijadikan referensi kedepan dalam meningkatkan program studi. Salah satu contohnya adalah data mahasiswa.

Menurut (Widhi, 2019) “pengolahan data berukuran besar yang mempunyai jumlah record dan jumlah atribut yang cukup banyak tidak dapat dilakukan dengan mudah. Salah satu cara yang dapat diterapkan untuk mengolah data yang berukuran besar adalah dengan menerapkan penggunaan *Data Mining*. Teknologi *data mining* merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi”. Karena data mining sendiri mempunyai peluang dan teknik untuk memenuhi kebutuhan, salah satunya adalah kebutuhan akan informasi yang komprehensif, dan informasi yang diperoleh dapat digunakan untuk pengambilan keputusan atau untuk menentukan kualitas pengambilan keputusan. Pengumpulan data atau informasi dapat dijadikan kesimpulan untuk pengambilan keputusan dengan cara

menganalisis dan menelaah informasi yang terkandung dalam data tersebut. (Widhi, 2019).

(Marajari, 2021) mengatakan bahwa “berdasarkan berlimpahnya data mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga berguna bagi pihak universitas”. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (*knowledge discovery*), misalnya informasi mengenai pengelompokan data mahasiswa berdasarkan daerah asal mahasiswa. Pengetahuan baru tersebut dapat membantu pihak universitas dalam meningkatkan kualitas mahasiswa dengan tepat sasaran. Salah satu langkah yang dapat dilakukan untuk meningkatkan kualitas mahasiswa adalah dengan memanfaatkan data mahasiswa diintegrasikan dengan teknik *data mining* sehingga menghasilkan kesimpulan (Marajari, 2021).

Menurut (Widhi, 2019) “*Data Mining* adalah proses untuk mendapatkan informasi dengan melakukan pencarian pola dan relasi-relasi yang tersembunyi di dalam timbunan data yang banyak. Data Mining atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran *Data Mining* ini bisa dipakai untuk membantu pengambilan keputusan di masa depan. Pengembangan KDD ini menyebabkan penggunaan *pattern recognition* semakin berkurang karena telah menjadi bagian *Data Mining*”.

Banyaknya mahasiswa prodi pendidikan matematika adalah lebih dari 200 mahasiswa dengan tingkatan semester yang berbeda. Data Mahasiswa Matematika Universitas Nias dalam penelitian ini adalah berisi data Kabupaten/kota Mahasiswa berasal, Jurusan yang diambil pada saat SMA/SMK, jenis kelamin, semester, usia, rata-rata durasi belajar mandiri per hari, jumlah Satuan Kredit Semester (SKS), nilai Indeks Prestasi Semester (IPS), dan nilai Indeks Prestasi Kumulatif (IPK) pada semester terakhir mahasiswa. “Variabel IPK, dan IPS adalah penilaian hasil belajar

yang diperoleh oleh mahasiswa setiap akhir semesternya dan dalam mencapai hasil terbaik pada penilaian tersebut perlu didukung oleh proses belajar yang tinggi, hal ini kemudian diambil peneliti yang didefinisikan sebagai Durasi Belajar. Sebagian besar mahasiswa melakukan kesalahan dalam hal durasi belajar. Banyak mahasiswa yang lupa bahwa untuk mendapatkan hasil yang besar diperlukan usaha yang besar pula. Oleh karena itu peneliti tertarik untuk mengelompokkan mahasiswa berdasarkan kategori rata-rata durasi belajar dan usia mahasiswa. Selanjutnya Peneliti juga menambahkan faktor lain sebagai variabel dalam penelitian ini yakni penerima beasiswa dan pendapatan orang tua”.

Tujuan pengolahan data mahasiswa program studi pendidikan matematika di Universitas Nias yaitu menerapkan algoritma *K-Means Clustering* pada data mahasiswa. Data yang digunakan adalah data mahasiswa semester 2 sampai dengan semester 8 dan sumber data berasal dari kuesioner yang telah disebar secara online. Manfaat yang diharapkan adalah sebagai referensi kedepannya bagi program studi dalam mengembangkan program studi matematika baik dari segi kuantitas maupun kualitas mahasiswanya.

Data yang terkumpul tidak dapat memberikan hasil yang maksimal jika tidak dianalisis dengan baik. Oleh karena itu, kita bisa menggunakan salah satu metode data mining yang umum digunakan yaitu *clustering*. Saat ini, analisis klaster mulai banyak digunakan di banyak bidang ilmu pengetahuan.

Dari pemaparan Wirnaning dan Kismiantini (2022), Analisis *cluster* ini merupakan salah satu teknik multivariat yang digunakan untuk mengklasifikasi objek atau kasus berupa responden ke dalam suatu kelompok yang relatif homogen, yang biasanya disebut dengan *cluster*. Jadi objek atau suatu kasus dalam setiap kelompok cenderung mirip satu sama lain dan berbeda jauh atau tidak sama dengan objek dari *cluster* lainnya. *Clustering* ini juga merupakan suatu metode untuk mengelompokkan dan mencari data dengan karakteristik yang memiliki kemiripan antara satu data dengan yang lainnya. Kelebihan dari analisis cluster ini adalah dapat

mengelompokkan data dalam jumlah besar dan variabel yang relatif banyak dan juga dapat dipakai dalam skala data ordinal, interval dan, rasio. Dalam *data mining* ada dua jenis metode *clustering* yang digunakan dalam mengelompokkan data, yaitu metode hierarki dan metode non-hierarki.

Uraian dari Anggi (2018) “salah satu metode non-hierarki yaitu metode *K-means* yang digolongkan sebagai metode pengklasifikasian yang bersifat *unsupervised* (tanpa arahan) dikarenakan data yang di analisis tidak mempunyai label kelas yang berarti dalam proses pengelompokannya, analisis ini tidak mempunyai anggota *cluster* yang pasti. Melainkan data yang telah masuk ke dalam *cluster* bisa saja berpindah ke *cluster* yang lain. Akan tetapi juga, karena peneliti sering menentukan sendiri jumlah cluster awal, baik dengan menggunakan metode tertentu atau berdasarkan pengalaman, maka metode *K-Means* ini sering disebut juga sebagai metode *semi-supervised classification*”. Pendapat Musfiani (2019) “*K-Means* ini juga merupakan salah satu metode data cluster non-hierarki yang berusaha mempartisi data ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi ke dalam cluster sehingga data yang memiliki karakteristik yang berbeda dikelompokkan pada kelompok lain”.

Berdasarkan uraian latar belakang sebelumnya, penulis tertarik untuk mengangkat judul penelitian ini “**Analisis Cluster Mahasiswa Pendidikan Matematika Universitas Nias**”.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah diuraikan diatas, dapat diidentifikasi masalah sebagai berikut ini:

- a. Banyaknya data khususnya program studi pendidikan matematika dan salah satu contoh datanya adalah data mahasiswa prodi pendidikan matematika.
- b. Memperoleh berbagai informasi data mahasiswa program studi pendidikan matematika yang dapat diolah.

1.3 Batasan Masalah

Agar materi dan pembahasan penelitian seperti yang diharapkan dan tidak meluas, maka penelitian ini diberi batasan. Beberapa batasan masalah pada penelitian sebagai berikut:

- a. Sasaran penelitian yakni mahasiswa program studi pendidikan matematika dari Universitas Nias dari semester 2-8.
- b. Penelitian ini menganalisis data menggunakan metode analisis cluster. Dimana ada banyak metode cluster maka dipilih salah satu yang lebih spesifik yaitu metode *K-means Cluster Analysis* dengan bantuan aplikasi IBM SPSS Statistic 26.

1.4 Rumusan Masalah

Berdasarkan uraian dari latar belakang dan identifikasi masalah yang telah dimuat diatas maka rumusan masalah dari penelitian ini adalah:

- a. Bagaimana penerapan metode *cluster analysis* dalam pembentukan cluster pada data mahasiswa program studi pendidikan matematika?
- b. Berapa cluster yang terbentuk dalam proses *cluster analysis*?
- c. Bagaimana hasil *cluster analysis* yang didapatkan?
- d. Bagaimana kesimpulan dari hasil *cluster analysis*?

1.5 Tujuan Penelitian

Berdasarkan paparan dari rumusan masalah di atas, maka tujuan penelitian ini adalah sebagai berikut:

- a. Mendeskripsikan data mahasiswa program studi pendidikan matematika di Universitas Nias
- b. Mengcluster data-data mahasiswa berdasarkan parameter yang ada
- c. Menampilkan hasil analisis untuk dapat digunakan sebagai referensi dalam meningkatkan kuantitas maupun kualitas dari mahasiswa program studi pendidikan matematika

1.6 Manfaat Penelitian

Adapun Manfaat dari penelitian adalah sebagai berikut:

1.6.1 Manfaat Teoritis

Dapat sebagai referensi dalam mengetahui *clustering* (pengelompokkan) mahasiswa prodi pendidikan matematika melalui proses analisis *cluster*.

1.6.2 Manfaat Praktis

- a. Untuk peneliti, manfaatnya yakni Ilmu yang diperoleh selama kuliah digunakan untuk mendukung persiapan transisi ke dunia kerja dan juga untuk menambah pandangan yang lebih luas tentang analisis klaster.
- b. Bagi Program Studi Matematika, dapat digunakan sebagai bahan penulisan ilmiah bagi pembaca dan sebagai referensi bagi mahasiswa terkait analisis cluster serta dapat digunakan sebagai bahan referensi untuk meningkatkan jumlah dan kualitas mahasiswa pada program pendidikan matematika.
- c. Bagi instansi, dapat dijadikan sebagai bahan penilaian kampus Universitas Nias untuk lebih memperhatikan daerah-daerah yang mempunyai potensi lebih besar untuk menarik mahasiswa baru matematika dan meningkatkan kualitas daerah yang lulusannya terutama daerah yang jumlah mahasiswa matematikanya terbanyak.

BAB II KAJIAN PUSTAKA

2.1 Kajian Teori

2.1.1 Analisis Cluster

a. Pengertian Cluster

Menurut (Dwi, 2021) “Analisis cluster merupakan suatu teknik yang digunakan untuk menghasilkan objek atau kasus (responden) ke dalam kelompok yang relatif homogen, yang disebut cluster.” Objek/kasus dalam setiap kelompok cenderung mirip atau sama satu sama lain dan berbeda jauh (tidak sama) dengan objek dari cluster lainnya.

Analisis cluster ini ⁸ pertama kali dilakukan oleh Tyron pada tahun 1939. Menurut (Khairun, 2019) “Analisis cluster bertujuan untuk mengalokasikan objek ke dalam suatu cluster yang independen sehingga objek-objek didalam satu cluster yang memiliki kesamaan antar objek serta objek-objek yang berada dalam cluster yang berbeda tidak mirip. Pada proses pengclusteran dilakukam satu ukuran yang dapat menghasilkan kesamaan atau kedekatan antar objek untuk mengetahui mirip atau berbedanya suatu data yang kompleks, yaitu ukuran jarak atau kemiripan.” Secara umum jarak yang digunakan adalah jarak euclidian (*euclidean distance*).

Menurut (Anggi, 2018) “tujuan utama dari analisis cluster adalah mengklasifikasi objek (kasus/elemen) seperti orang, produk (barang), toko, perusahaan, kedalam kelompok-kelompok yang relatif homogen didasarkan pada suatu set variabel yang dipertimbangkan untuk diteliti”.

¹ Analisis *cluster* membagi kumpulan data menjadi satu atau lebih cluster tertentu. Menurut (Rindang, 2021), *cluster* yang baik adalah *cluster* yang mempunyai:

1) Homogenitas atau kemiripan yang tinggi antara anggota dalam cluster (*within-cluster*).

- 2) Heterogenitas atau perbedaan besar antara satu cluster dengan cluster lainnya (*between cluster*).

Berdasarkan uraian di atas dapat disimpulkan bahwa cluster yang baik adalah cluster yang anggota-anggotanya semirip mungkin satu sama lain, namun tidak terlalu mirip dengan anggota cluster lainnya. Di sini kata 'mirip' diartikan sebagai tingkat kemiripan fitur antara dua kumpulan data.

b. ¹ Asumsi Pada Analisis Cluster

Menurut (Anggi, 2018) asumsi untuk analisis cluster yakni :

- 1) “Sampel yang diambil benar-benar mewakili populasi yang ada. Memang tidak ada ketentuan jumlah sampel yang representatif, namun tetaplah diperlukan sejumlah sampel yang cukup besar agar proses clustering bisa dilakukan dengan benar”.
- 2) **Multikolinearitas, yakni** keadaan dimana satu atau lebih variabel independen berkorelasi dengan variabel independen lainnya. Menurut (Hasiara, 2019), “uji multikolinieritas adalah suatu keadaan dimana satu atau lebih variabel bebas terdapat korelasi dengan variabel bebas lainnya. Uji multikolinieritas bertujuan untuk mengetahui model regresi berkorelasi antar variabel bebas atau independen. Uji Multikolinieritas dapat dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF) dari masing-masing variabel bebas terhadap variabel terikatnya. Jika nilai VIF tidak lebih dari 10, maka model dikatakan tidak terdapat gejala multikolinier”.

c. Proses Analisis Cluster

Berdasarkan pemaparan (Awaliah, 2018) dan juga (Dwi, 2021), analisis cluster terdiri dari beberapa proses dasar, yaitu:

- 1) Merumuskan masalah
Saat merumuskan masalah analisis cluster, pilihan variabel yang digunakan untuk clustering (pembentukan cluster) adalah

yang paling penting. Memasukkan satu atau dua variabel yang tidak terkait dengan masalah pengelompokan/pengelompokan akan mempengaruhi hasil pengelompokan, yang kemungkinan besar akan sangat berguna.

2) Memilih ukuran jarak atau similaritas

Karena tujuan clustering adalah untuk mengelompokkan objek-objek serupa ke dalam cluster yang sama, beberapa tindakan diperlukan untuk mengakses objek-objek yang serupa atau berbeda. Jadi pendekatan yang paling umum adalah mengukur kemiripan, yang dinyatakan sebagai jarak antar pasangan objek. Penghitungan jarak biasanya menggunakan jarak *Euclidean*, yang cocok untuk kumpulan data kecil. Benda-benda yang jaraknya lebih pendek mempunyai kemiripan atau lebih mirip satu sama lain dibandingkan berpasangan dengan jarak yang lebih jauh

3) Standarisasi data

Apabila terdapat perbedaan besar ukuran satuan antar variabel yang diteliti maka dilakukan proses standarisasi. Perbedaan satuan pengukuran yang signifikan dapat menyebabkan kesalahan perhitungan dalam analisis klaster. Oleh karena itu perlu dilakukan proses standarisasi dengan transformasi (standarisasi data mentah sebelum dilakukan analisis lebih lanjut. Transformasi dilakukan dengan variabel yang sesuai berupa Z-score, seperti berikut :

$$Z = \frac{x - \bar{x}}{\sigma} \quad (2.1)$$

Keterangan:

x = nilai data

\bar{x} = nilai rata-rata

σ = standar deviasi

4) Memilih suatu prosedur pengclusteran

Menurut (Anggi, 2018), setelah melakukan standarisasi data yang dianggap mengandung satuan yang sangat berbeda, dan menentukan metode cluster, langkah selanjutnya adalah mengcluster data.

5) Melakukan interpretasi terhadap cluster yang telah terbentuk

Menurut (Dwi, 2019) “setelah jumlah cluster terbentuk dengan metode hirarki atau non-hirarki, langkah selanjutnya adalah melakukan interpretasi terhadap cluster yang telah terbentuk. Tahap interpretasi meliputi pengujian tiap cluster dalam terminologi untuk menamai dan menandai dengan suatu label yang secara akurat dapat menjelaskan kealamian cluster. Proses ini dimulai dengan suatu ukuran yang sering digunakan yaitu *Centroid Cluster*. Membuat profil dan interpretasi cluster tidak hanya untuk memperoleh suatu gambaran saja, melainkan untuk menyediakan rata-rata untuk menilai korespondensi pada cluster yang terbentuk serta profil cluster memberikan arahan bagi signifikansi praktis”.

d. Metode Analisis Cluster

Ada dua metode analisis klaster dalam analisis klaster ini yang dapat digunakan yakni menurut (Dwi, 2019) “metode hierarki dan metode non-hierarki. Keduanya memiliki kelebihan dan kekurangan masing-masing. Kelebihan metode hierarki yaitu cepat dalam proses pengolahan sehingga menghemat waktu, namun kekurangannya yaitu metode ini dapat menimbulkan kesalahan. Selain itu tidak baik diterapkan untuk menganalisis sampel dengan ukuran besar. Metode non-hierarki memiliki kelebihan yang lebih baik dibandingkan metode hierarki. Hasilnya memiliki sedikit kelemahan pada data outlier”. Data outlier adalah suatu data yang memunculkan nilai-nilai yang berlebihan atau melebihi batas dan jauh berbeda dengan data-data yang masih masuk dalam satu subset data. Ukuran jarak yang

digunakan dan termasuk variabel tak relevan atau variabel yang tidak tepat (Anggi, 2018).

Keuntungannya adalah bahwa hanya menggunakan titik awal non-acak, metode non-hierarki jauh lebih buruk daripada metode hierarki untuk titik awal acak. Penentuan metode mana yang akan digunakan bergantung pada peneliti dan konteks penelitian, apapun topik, teori, dan konsep yang diterapkan. Pilihan lainnya adalah menggabungkan kedua metode tersebut. Gunakan metode hierarki terlebih dahulu, lalu lanjutkan ke metode non-hierarki.

a. Metode Hierarki

Metode hierarki (*hierarchical method*) merupakan metode yang dilakukan secara bertahap. Metode ini membentuk tahapan tertentu sebagai struktur pohon dan dapat membuat dendogram. Dendogram adalah diagram bercabang-cabang menyerupai pohon yang dipakai untuk menggambarkan derajat kekerabatan atau kesamaan. Dalam (Musfiani, 2019) menguraikan juga bahwa dalam metode hierarki, pengelompokan dimulai dengan dua atau lebih objek yang mempunyai kesamaan yang paling dekat.

Pengelompokan dilanjutkan ke objek lainnya yang mempunyai jarak yang dekat dan demikian seterusnya hingga objek tergabung dalam cluster sehingga membentuk hierarki yang jelas seperti pohon (dendogram). Menurut (Khairun, 2019) “dendrogram merupakan representasi visual tahapan proses analisis massa dan nilai koefisien jarak untuk setiap tahapan. Hasil numerik di sebelah kanan dendogram merupakan objek yang akan diteliti karena terdapat garis-garis yang menghubungkan objek tersebut dengan objek lain dalam satu cluster”. Berikut ini diberikan ukuran ketidakmiripan antar cluster (Khairun, 2019) :

1) *Pautan Tunggal (Single Linkage/Nearest Neighbour Method)*

Metode ini akan mengelompokkan dua objek yang mempunyai jarak terdekat terlebih dahulu.

2) *Pautan Lengkap (Complete Linkage/Furthest Neighbour Method)*

Metode ini akan mengelompokkan dua objek yang mempunyai jarak terjauh terlebih dahulu.

3) *Pautan Rataan (Average Linkage/Between Neighbour Method)*

Menurut (Anggi, 2018) “metode ini akan mengelompokkan objek berdasar jarak rata-rata yang didapat dengan melakukan rata-rata semua jarak antar objek terlebih dahulu. Hasil berupa *Average Linkage Clustering* dapat disajikan dalam bentuk suatu dendogram atau diagram pohon. Cabang-cabang pohon menunjukkan cluster (kelompok). Cabang-cabang tersebut bertemu bersama-sama pada simpul yang posisinya sepanjang suatu sumbu jarak (kemiripan)”.

4) *Ward's Error Sum of Square Method*

Menurut (Anggi, 2018) “pada metode ini, jarak antara dua cluster yang terbentuk adalah sum of squares diantara dua cluster tersebut. Metode Ward berusaha untuk meminimalkan variasi antar objek dalam satu cluster dan memaksimalkan variasi dengan objek yang ada di cluster lainnya. Metode Ward didasarkan pada kriteria *sum square error* (SSE) dengan ukuran kehomogenan antara dua objek berdasarkan jumlah kuadrat kesalahan minimal”.

5) *Metode Centroid*

Menurut (Dwi, 2019) “*Centroid method* disebut juga metode titik pusat, dimana jarak antar cluster pada metode centroid merupakan jarak antar centroid. Jika terjadi

pembentukan cluster baru maka akan terjadi perhitungan ulang”

8
b.

Metode Non-Hierarki

Menurut (Khairun, 2019) “metode non-hierarki disebut juga dengan metode *k-means*. Metode ini tidak sama halnya dengan metode hierarki, karena pada metode non hierarki dimulai dengan menentukan terlebih dahulu sejumlah cluster awal yang diinginkan, selanjutnya hasil dari objek pengamatan tersebut bergabung dan membentuk cluster”. “Keuntungan penggunaan metode non-hierarki dapat melakukan analisis sampel dalam ukuran yang lebih besar dan lebih efisien. Selain itu, hanya memiliki sedikit pada data outlier, ukuran jarak yang digunakan, dan variabel tak relevan atau variabel yang tidak tepat. Sedangkan kelemahannya adalah untuk titik bakal random lebih buruk dari pada metode hierarki” (Dwi, 2021).

Dalam (Musfiani, 2019), metode non-hierarki memiliki tiga prosedur yaitu :

1) *Sequential Threshold*

“Proses clustering dengan menggunakan metode *Sequential Threshold* dimulai dengan memilih sebuah *cluster seed* dan kemudian menggabungkan setiap objek yang ada dalam jarak yang telah ditentukan sebelumnya. Cluster tersebut akan disebut sebagai cluster pertama. Setelah cluster pertama terbentuk, maka *cluster seed* kedua digunakan dan kemudian objek-objek yang mempunyai jarak terdekat akan digabungkan. Jika telah selesai, maka cluster berikutnya akan dibentuk dengan cara yang sama” (Anggi, 2018).

Cara ini disebut dengan *sequential Thresholding* karena proses clustering dilakukan secara berurutan mulai dari cluster pertama, kedua, ketiga, dan berikutnya. Cara ini tidak memperbolehkan objek berpindah ke cluster lain.

2) *Parallel Threshold*

“Proses clustering dengan menggunakan metode *Parallel Threshold* dimulai dengan memiliki cluster seed yang akan dijadikan patokan pembuatan cluster. Setiap objek akan diukur terhadap cluster seed tersebut. Sebuah objek akan masuk ke suatu cluster jika mempunyai nilai jarak terhadap suatu cluster seed lebih dekat daripada cluster seed yang lain. Langkah ini mengakibatkan penentuan cluster tidak berurutan” (Anggi, 2018)..

Metode ini disebut *Parallel Threshold* karena proses clustering dilakukan tidak berurutan. Metode ini tidak memperbolehkan suatu objek yang sudah menjadi anggota suatu cluster berpindah ke cluster lain.

3) *Optimizing Partitionin (Metode Partisi)*

Proses clustering dengan menggunakan metode ini mirip dengan *Sequential Threshold* atau *Parallel Threshold clustering*, Namun yang membedakan pada metode ini memungkinkan untuk menempatkan kembali hasil objek-objek ke dalam cluster yang lebih dekat (Khairun, 2019).

Proses *Optimizing Partitionin* menurut (Khairun, 2019) “dilakukan dengan menghitung ulang setiap objek terhadap semua *centroid cluster* yang ada. Jika objek tersebut mempunyai jarak yang lebih kecil terhadap suatu *centroid cluster* lain daripada *centroid cluster* sekarang, maka objek tersebut di relokasi ke cluster terdekat tersebut”.

2.1.2 *K-Means Cluster*

a. *Pengertian K-Means Cluster*

Menurut (Ahmad et al., 2022) “*K-Means Cluster* adalah salah satu metode dan *clustering* non-hierarki yang berusaha mengelompokkan data ke dalam suatu cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama. *K-Means* termasuk algoritma *clustering* dengan proses berulang-ulang. Huruf K diartikan sebagai jumlah cluster yang hendak dibuat. Selanjutnya nilai K ditetapkan secara acak. Sedangkan *Means* adalah nilai sementara yang menjadi pusat dari cluster atau disebut juga dengan *centroid*. Setiap data yang ada dihitung jaraknya terhadap masing-masing *centroid* dengan memakai rumus *Euclidean* hingga dihasilkan jarak terdekat dari setiap data dengan *centroid*”.

Menurut (Anggi, 2018) “berbeda dengan metode hierarki, metode non-hierarki justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau yang lain). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means cluster*”. Menurut (D. N. P. Sari & Sukestiyarno, 2021) “metode *K-means* berusaha mengelompokkan data yang ada ke dalam satu kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.”

Metode *K-Means* adalah metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada numerik (Rindang, 2021; Wirnaning et al., 2022).

b. Langkah-langkah Analisis *K-Means Cluster*

(Rindang, 2021) menjelaskan “data clustering menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:”

- 1) Menentukan jumlah *cluster*

Langkah pertama dalam analisis cluster menggunakan Metode K adalah menentukan ukuran k, atau jumlah cluster, dan menentukan mean setiap cluster, atau pusat cluster. Penentuan jumlah cluster dan sentrida pada setiap cluster bersifat objektif..

2) Melakukan standarisasi data

Yakni konversi ke format Z-Score ketika terdapat perbedaan besar ukuran unit antara data dan variabel. Standardisasi menggunakan rumus Z-Score yaitu:

$$Z = \frac{x - \bar{x}}{\sigma} \quad (2.2)$$

Keterangan:

Z = Nilai Z-Score

x = nilai data

\bar{x} = nilai rata-rata

σ = standar deviasi

3) Menentukan nilai centroid untuk awal iterasi secara acak.

Selanjutnya jika menentukan nilai centroid yang merupakan tahap dari iterasi, dapat digunakan rumus:

$$\bar{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2.3)$$

Dimana:

\bar{V}_{ij} = centroid/rata-rata cluster ke-i untuk variable ke-j

i, k = indeks dari cluster j adalah indeks dari variable

N_i = jumlah data yang menjadi anggota cluster ke-i

x_{kj} = nilai data ke-k yang ada di dalam cluster tersebut untuk variable ke-j

4) Menghitung Euclidean Distance untuk mengetahui jarak antara titik centroid dengan titik tiap objek menggunakan yaitu:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2.4)$$

Dimana:

D_e = Euclidean Distance

i = banyaknya objek

²
(s, t) = koordinat centroid

(x, y) = koordinat objek

- 5) Pengelompokkan objek dalam menentukan anggota *cluster* dengan menghitung jarak minimum objek. Nilai yang diperoleh adalah 0 atau 1, nilai 1 berarti data yang dialokasikan ke *cluster* dan nilai 0 data yang dialokasikan ke *cluster* yang lain.
- 6) Kembali ke langkah 2, ulangi hingga nilai centroid tetap tercapai dan anggota cluster tidak berpindah ke cluster lain.

2.1.3 SPSS

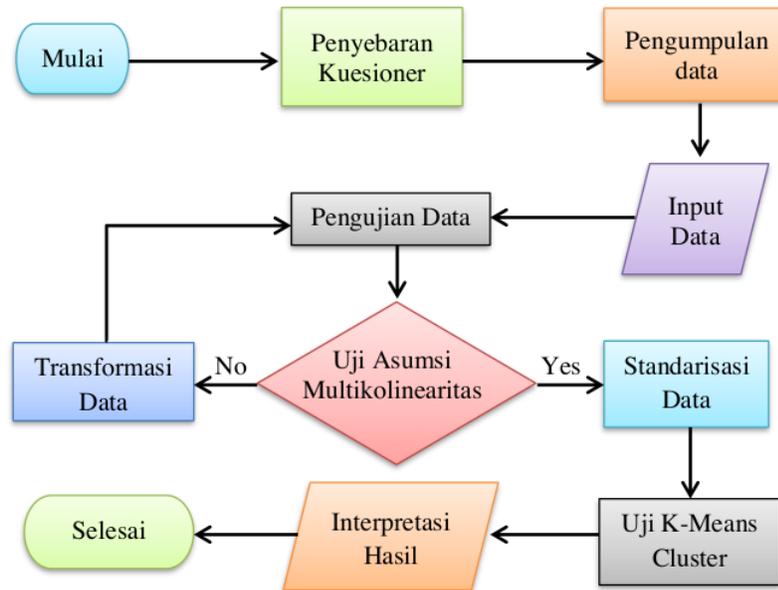
Menurut (Ahmad et al., 2022) “problem yang dihadapi oleh para pengguna statistik umumnya adalah penyelesaian yang menyita banyak waktu, diperlukan ketelitian, kecermatan, mampu menghafal, mengingat dan mengaplikasikan rumus-rumus statistik, sehingga statistik menjadi ilmu yang dekat dengan persepsi ‘rumit’. Seiring dengan perkembangan teknologi informasi yang cepat, kerja-kerja statistik saat ini bisa diselesaikan dengan perangkat komputer dengan aplikasi (software) statistik. Penggunaan teknologi ini sangat membantu penggunaan statistik dengan cepat, cermat dan akurat menyelesaikan persoalan-persoalan data statistik. Sehingga aktivitas kerja menjadi efektif dan efisien. Salah satunya adalah SPSS”.

Menurut Priyatno dalam Anggi (2018) “SPSS merupakan program untuk olah data statistik yang paling populer dan paling banyak pemakaiannya di seluruh dunia dan banyak digunakan oleh para peneliti untuk berbagai keperluan seperti riset pasar, untuk menyelesaikan tugas penelitian seperti skripsi, tesis, disertasi dan sebagainya”.

Dengan SPSS memudahkan penelitian dan analisis data. Menurut (Ahmad et al., 2022) “SPSS mempunyai banyak kelebihan dibandingkan dengan program aplikasi statistik lainnya, diantaranya mudah dioperasikan, mampu mengakses data dari berbagai macam format data, serta mampu menampilkan data yang lebih informatif dan informasi yang akurat”

2.2 Kerangka Berpikir

Kerangka berpikir dalam penelitian ini dapat digambarkan sebagai berikut.



Gambar 2.1 Langkah-Langkah Penelitian

BAB III METODE PENELITIAN

3.1 Jenis Penelitian

Pada penelitian ini, jenis penelitian yang digunakan ialah penelitian terapan (*Applied Research*) dengan pendekatan kuantitatif. Penelitian terapan bertujuan untuk menjawab permasalahan khusus atau untuk membuat keputusan tentang suatu kebijakan. Penggunaan metode ilmiah dalam penelitian terapan menjamin objektivitas dalam pengumpulan fakta dan menguji kreatifitas. Data yang digunakan dalam penelitian ini adalah data primer yang diperoleh langsung dari responden dengan menggunakan kuesioner dan kemudian ditabulasi. Hasil tabulasi data diolah dengan SPSS.

Metode penelitian ini merupakan proses kegiatan dalam bentuk pengumpulan data mahasiswa program studi pendidikan matematika, menganalisis data dan memberikan interpretasi yang terkait dengan tujuan penelitian.

3.2 Variabel Penelitian

Adapun variabel yang digunakan pada penelitian ini adalah sebagai berikut:

- X₁ = Kabupaten/kota Mahasiswa berasal
- X₂ = Asal Sekolah Mahasiswa
- X₃ = Usia mahasiswa
- X₄ = Semester mahasiswa
- X₅ = Rata – rata durasi belajar mandiri mahasiswa per hari (Jam)
- X₆ = Jumlah SKS mahasiswa pada semester terakhir
- X₇ = IPS mahasiswa pada semester terakhir
- X₈ = IPK mahasiswa pada semester terakhir
- X₉ = Memiliki Laptop
- X₁₀ = Tempat tinggal mahasiswa sekarang
- X₁₁ = Memiliki Kendaraan (Sepeda motor)

X₁₂ = Penerima Beasiswa

X₁₃ = Pendapatan Orang Tua Mahasiswa

3.3 Subjek Penelitian

Dalam penelitian ini yang menjadi subjek penelitian ini adalah seluruh mahasiswa program studi pendidikan matematika di Universitas Nias dari semester 2 sampai dengan semester 8.

3.4 Instrumen Penelitian

Instrumen penelitian dalam penelitian ini adalah kuesioner yang dibagikan dan diisi oleh subjek. Menurut Hasiara et all (2019) Kuesioner merupakan suatu teknik pengumpulan data yang memungkinkan responden memperoleh informasi melalui kuesioner. Kuesioner tertutup memberikan alternatif jawaban terhadap setiap pertanyaan, sehingga responden hanya tinggal memilih di antara jawaban yang tersedia. Pada saat yang sama, kuesioner terbuka memungkinkan responden untuk menyajikan jawaban mereka secara objektif. Adapun kisi-kisi instrumen adalah sebagai berikut.

Tabel 3.1 Kisi-kisi Instrumen untuk Mahasiswa:

Variabel	Indikator	Nomor Pertanyaan
Data Pribadi Responden	Nama	A1
	NIM	A2
	Jenis Kelamin	A3
	Usia	A4
	Tempat, Tanggal Lahir	A5
	Alamat	A6, A7, A8
	Kontak	A9, A10
Riwayat Sekolah Asal Responden	Nama Sekolah	B1
	Jurusan	B2
	Tahun Tamat	B3
Riwayat Perkuliahan Responden	Semester	C1
	Rata-Rata Belajar	C2
	Jumlah SKS yang diambil	C3

semester terakhir	IPS	C4
	IPK	C5
	Peralatan Elektronik	C6
	Jarak Rumah	C7
	Tempat Tinggal	C8
	Kendaraan	C9
Tingkat Ekonomi	Penerima Beasiswa	D1
	Pekerjaan Orang Tua	D2
	Pendapatan Orang Tua	D3

3.5 Teknik Pengumpulan Data

Teknik pengumpulan data yang akan dipakai dalam penelitian ini adalah berupa:

a. Dokumentasi

Dokumentasi merupakan teknik pengumpulan data dengan cara memperoleh informasi dari bermacam-macam sumber tertulis atau dokumen yang ada pada responden atau tempat, dimana responden bertempat tinggal atau melakukan kegiatan sehari-harinya. Pada penelitian ini, peneliti menggunakan dokumentasi sebagai sarana untuk mendapatkan data yang diinginkan. Dokumentasi dalam penelitian ini berupa data Kabupaten/kota Mahasiswa berasal, Jurusan yang diambil pada saat SMA/SMK, jenis kelamin, semester, umur, rata-rata lama belajar mandiri per hari, jumlah SKS, nilai IPS, dan nilai IPK pada semester ganjil Tahun Akademik 2022/2023, serta penerima beasiswa atau bukan.

b. Kuesioner

Kuesioner merupakan teknik pengumpulan data yang merupakan satu set pertanyaan yang berurusan dengan satu topic tertentu yang saling berkaitan, yang harus dijawab oleh subjek. Kuesioner ini digunakan untuk penyelidikan mengenai suatu topic tertentu, dengan cara mengedarkan formulir daftar pertanyaan tersebut agar dijawab atau direspon oleh subjek penelitian (Zulfah,

2022). Kuesioner dalam penelitian ini merupakan kumpulan pertanyaan berupa data Kabupaten/kota Mahasiswa berasal, Jurusan yang diambil pada saat SMA/SMK, jenis kelamin, semester, umur, rata-rata lama belajar mandiri per hari, jumlah SKS, nilai IPS, dan nilai IPK pada semester ganjil Tahun Akademik 2022/2023, serta penerima beasiswa atau bukan.

3.6 Teknik Analisis Data

Analisis data yang digunakan adalah sebagai berikut:

1. Mengumpulkan instrumen yang telah dibagikan dan diisi oleh responden kemudian diperiksa kelengkapannya.
2. Gunakan SPSS Data View dan Variable View untuk mengatur semua variabel sebagai aturan utama sistem data numerik.
3. Masukkan informasi instrumen pada tabel data penelitian.
4. Melakukan uji asumsi multikolinearitas terhadap data.
5. Melakukan standarisasi data.
6. Menentukan prosedur analisis cluster.
Penelitian ini menggunakan metode analisis cluster nonhierarki dengan menggunakan *K-Means Cluster*.
7. Jalankan hasil analisis cluster.
8. Setelah menerima hasil analisis klaster, langkah selanjutnya adalah menginterpretasikan hasil klaster yang terbentuk.
9. Kesimpulan dan saran.

3.7 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Universitas Nias yang terletak di Jln. Yos. Sudarso No. 118/E-S, Ombolata Ulu, Kota Gunungsitoli, Sumatera Utara, Nias 22812.

BAB IV HASIL DAN PEMBAHASAN

1.1 Hasil Penelitian

1.1.1 Data

Menurut Kamus Besar Bahasa Indonesia, data ialah keterangan yang benar dan nyata, keterangan atau bahan nyata yang dapat dijadikan dasar kajian atau informasi dalam bentuk yang dapat diproses oleh computer. Dalam penelitian ini, sebelum proses perhitungan data dilakukan pengumpulan data terlebih dahulu. Pengumpulan data dilakukan menggunakan *Google Form* dengan menyebarkan kuesioner kepada siswa sebagai responden melalui grup *WhatsApp* atau menghubungi responden secara langsung. Perangkat lunak yang digunakan untuk pengolahan data adalah *IBM SPSS Statistics 26*. Informasi yang diperoleh ada pada Lampiran 1 dan jumlah responden sebanyak 203 responden mahasiswa pendidikan matematika.

1.1.2 Inisialisasi Data

Pada Lampiran 1, kolom Kabupaten, Asal Sekolah, Jenis Kelamin, Durasi Belajar, Memiliki Laptop, Tempat Tinggal, Memiliki Kendaraan, Beasiswa, Pendapatan Orangtua berupa data kualitatif, karena semua data mempunyai status yang sama dalam arti tidak ada data yang mempunyai tingkat lebih tinggi atau lebih rendah dibandingkan dengan data lainnya, oleh karena itu perlu dilakukan inisialisasi terlebih dahulu data tersebut ke dalam bentuk numerik. Adapun data yang diinisialisasi peneliti ditunjukkan dalam Tabel berikut.

Tabel 4.1 Tabel inisialisasi data

No	Variabel	Inisial	
1	Kabupaten	Kabupaten Nias	1
		Kabupaten Nias Barat	2
		Kabupaten Nias Selatan	3
		Kabupaten Nias Utara	4

		Kota Gunungsitoli	5
		Yang Lain: Kabupaten Tapanuli Tengah	6
2	Asal Sekolah	SMA	1
		SMK	2
		MA	3
3	Durasi Belajar	Kurang dari 1 jam	1
		1 - 2 jam	2
		2 - 3 jam	3
		3 - 4 jam	4
		> 4 Jam	5
4	Memiliki Laptop	Ya	1
		Tidak	2
5	Tempat Tinggal	Kos	1
		Rumah	2
		Asrama	3
		Tempat Saudara	4
6	Memiliki Kendaraan	Ya	1
		Tidak	2
7	Beasiswa	Ya	1
		Tidak	2
8	Pendapatan Orangtua	Kurang dari Rp. 500.000	1
		Rp.500.000 – Rp. 1.000.000	2
		Rp. 1.000.000 – Rp. 2.000.000	3
		Diatas Rp. 2.000.000	4

Data tersebut kemudian disusun kembali berdasarkan Tabel 4.1, dengan definisi variabel operasional sebagai berikut:

X₁ = Kabupaten/kota Mahasiswa berasal

X₂ = Asal Sekolah Mahasiswa

X₃ = Usia mahasiswa

X₄ = Semester mahasiswa

- X₅ = Rata – rata lama belajar mandiri mahasiswa per hari (Jam)
 - X₆ = Jumlah SKS mahasiswa pada semester terakhir
 - X₇ = IPS mahasiswa pada semester terakhir
 - X₈ = IPK mahasiswa pada semester terakhir
 - X₉ = Memiliki Laptop
 - X₁₀ = Tempat tinggal mahasiswa sekarang
 - X₁₁ = Memiliki Kendaraan (Sepeda motor)
 - X₁₂ = Penerima Beasiswa
 - X₁₃ = Pendapatan Orang Tua Mahasiswa
- Sehingga diperoleh data baru yang sudah diinisialisasi yang dapat dilihat pada Lampiran 2.

1.1.3 Uji Asumsi Multikolinearitas

Dalam (Desy, 2021), “Uji multikolinearitas dilakukan untuk menguji ada atau tidaknya variabel bebas yang mempunyai kemiripan antara variabel bebas lain. Besaran (*quality*) yang digunakan untuk mendeteksi adanya multikolinearitas adalah factor inflasi ragam (*Variance Inflation Factor/VIF*)”. Jika data menunjukkan multikolinearitas, salah satu opsi yang dapat dilakukan adalah dengan melakukan transformasi data. Cara untuk menguji adanya multikolinieritas, yaitu:

- a. Melihat nilai *tolerance*
Terjadi multikolinieritas, jika nilai *tolerance* $\leq 0,1$.
- b. Melihat nilai VIF (*Variance Inflation Factor*)
Terjadi multikolinieritas, jika nilai VIF ≥ 10 .

Berdasarkan Lampiran 2, dilakukan uji asumsi multikolinearitas dengan hasil sebagai berikut.

Tabel 4.2 Hasil Uji Multikolinearitas

Model		Collinearity Statistics	
		Tolerance	VIF
1	X1	.811	1.233
	X2	.887	1.128

X3	.450	2.223
X4	.239	4.182
X5	.923	1.084
X6	.167	5.991
X7	.103	9.680
X8	.275	3.639
X9	.816	1.226
X10	.772	1.295
X11	.814	1.228
X12	.844	1.185
X13	.828	1.208

Dari Tabel 4.2 terlihat bahwa seluruh variabel penelitian tidak mengalami multikolinearitas karena nilai toleransinya lebih besar dari 0,1 dan nilai VIF kurang dari 10, sehingga proses clustering dapat dilanjutkan apabila memenuhi asumsi non-multikolinearitas.

1.1.4 Standarisasi Data

Standarisasi data sangat diperlukan untuk standarisasi data guna meminimalkan pengaruh perbedaan skala atau satuan. Untuk melakukan ini, data dikurangi dari mean dan dibagi dengan standar deviasi. Standarisasi data dilakukan dengan mengkonversi data ke format z-score menggunakan rumus standarisasi data atau menggunakan IBM SPSS Statistics 26. Hal ini dapat dilihat pada Lampiran 3. Berikut statistic deskriptif sebagai dasar perhitungan *z-score*.

Tabel 4.3 Statistik deskriptif dasar perhitungan *z-score*

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
X1	203	1	6	3.64	1.468
X2	203	1	3	1.33	.494
X3	203	18	25	20.87	1.444
X4	203	2	8	4.92	2.136
X5	203	1	5	2.27	1.019
X6	203	11	24	20.67	4.393
X7	203	1.54	3.83	2.8479	.63983
X8	203	2.16	3.75	3.1105	.26663

X9	203	1	2	1.29	.455
X10	203	1	4	1.68	.697
X11	203	1	2	1.52	.501
X12	203	1	2	1.80	.399
X13	203	1	4	1.92	.989
Valid N (listwise)	203				

1.1.5 Asumsi Dalam Analisis Cluster

Sebelum melakukan clustering, harus dipenuhi dua prasyarat yaitu sampel representatif dan non-multikolinearitas.

a. Sampel Mewakili Populasi

Untuk mengetahui sampel mewakili populasi dapat dilihat dengan cara melihat nilai KMO (Jika nilai KMO > 0,5 disimpulkan mewakili, dan jika KMO < 0,5 disimpulkan tidak mewakili) (Ahmad, 2022). Nilai KMO dari data Lampiran 3 ditampilkan dalam tabel berikut.

Tabel 4.4 Nilai KMO

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.624
Bartlett's Test of Sphericity	Approx. Chi-Square	971.870
	df	78
	Sig.	.000

Berdasarkan Tabel 4.4 nilai KMO sebesar 0,624 > 0,5, sehingga dapat disimpulkan bahwa sampel tersebut mewakili populasi umum atau merupakan sampel yang representatif.

b. Nonmultikolinearitas

Berdasarkan hasil uji multikolinearitas pada Tabel 4.2 terlihat bahwa seluruh variabel penelitian tidak mengalami multikolinearitas karena nilai toleransinya lebih besar dari 0,1 dan nilai VIF kurang dari 10, sehingga asumsi non multikolinearitas terpenuhi.

1.1.6 Proses Clustering

a. Pembentukan *K-Means Cluster*

Analisis cluster dengan metode *K-means Cluster* dilakukan dengan bantuan IBM SPSS Statistic 26. Pertama ditentukan jumlah *cluster* adalah 4. Hasil Output K-Means Cluster dari SPSS diuraikan sebagai berikut.

1) *Initial Cluster*

Tabel 4.5 Centroid Awal Cluster

	Cluster			
	1	2	3	4
Zscore: X1	-1.79588	-1.79588	.24840	.92983
Zscore: X2	-.67859	-.67859	-.67859	-.67859
Zscore: X3	.78443	2.16913	.09209	-1.29261
Zscore: X4	1.44374	.50738	-1.36532	-1.36532
Zscore: X5	1.70212	-.26112	-1.24274	.72050
Zscore: X6	-1.97464	.75689	.30163	.30163
Zscore: X7	-1.84092	1.23801	-.52808	1.33178
Zscore: X8	-.41460	.86060	-2.25238	2.21081
Zscore: X9	-.63852	-.63852	1.55841	1.55841
Zscore: X10	-.97506	3.32792	1.89359	.45927
Zscore: X11	-1.03255	-1.03255	.96371	-1.03255
Zscore: X12	.49416	-2.01368	.49416	.49416
Zscore: X13	.08468	.08468	.08468	.08468

Tabel 4.5 diatas merupakan tampilan awal proses clustering sebelum dilakukan proses iterasi.

2) *Iterasi Analisis Cluster*

Dilakukan proses iterasi yang dilakukan pada objek sebanyak 10 sampel, dapat dilihat pada tabel.

Tabel 4.6 Iterasi

Iteration	Change in Cluster Centers			
	1	2	3	4
1	2.546	2.965	3.268	3.005

2	.559	.763	.467	.342
3	.085	.372	.165	.244
4	.000	.223	.074	.124
5	.000	.241	.000	.108
6	.000	.000	.046	.047
7	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 7. The minimum distance between initial centers is 5,971.

Tabel 4.6 menunjukkan bahwa proses iterasi dilakukan sebanyak 7 kali. Proses ini dilakukan untuk mendapatkan cluster yang tepat. Terlihat jarak minimum antar pusat cluster yang dihasilkan dari hasil iterasi adalah 5,971.

3) Jarak antar cluster

Tabel 4.7 Jarak Antar Cluster

Distances between Final Cluster Centers				
Cluster	1	2	3	4
1		4.655	4.157	4.253
2	4.655		3.607	2.664
3	4.157	3.607		2.512
4	4.253	2.664	2.512	

Distance-between atau *centroid-distance* adalah jarak antar *centroid* dari satu *cluster* ke *cluster* lainnya. Hasil Cluster dianggap baik jika mempunyai nilai jarak yang relatif tinggi. Nilai yang lebih besar berarti jarak antara cluster yang satu dengan cluster yang lain semakin jauh.

Berdasarkan tabel 4.7, jarak antar cluster 1 dan cluster 2 yaitu 4.655, jarak antar cluster 1 dan cluster 3 yaitu 4.157, jarak antar cluster 1 dan cluster 4 yaitu 4.253, jarak antar cluster 2 dan cluster 3 yaitu 3.607, jarak antar cluster 2 dan cluster 4 yaitu 2.664, serta jarak antar cluster 3 dan cluster 4 yaitu 2.512.

Artinya, perbedaan antara klaster yang satu dengan klaster lainnya semakin terlihat jelas. Dari hasil olah data menggunakan IBM SPSS Statistic 26 didapatkan jarak antar cluster relatif tinggi sehingga dapat dikatakan hasil dari clustering baik.

4) Jumlah anggota cluster

Berdasarkan *output Number of Cases in each Cluster*, jumlah anggota masing-masing cluster disajikan pada tabel berikut.

Tabel 4.8 Jumlah Anggota Cluster

Number of Cases in each Cluster		
Cluster		
	1	45.000
	2	29.000
	3	65.000
	4	64.000
Valid		203.000
Missing		.000

Dari 203 data mahasiswa, dibentuk 4 buah *cluster* dimana *cluster* 3 merupakan *cluster* dengan jumlah terbanyak yaitu 65 data mahasiswa, *cluster* 1 dengan 45 data mahasiswa, *cluster* 2 dengan 29 data mahasiswa serta *cluster* 4 dengan jumlah 64 data mahasiswa. Rincian anggota tiap *cluster* dapat dilihat pada Lampiran 4 *Cluster Membership*.

5) Pengujian tingkat signifikan

Untuk melihat uji tingkat signifikansi antar cluster dan mengetahui perbedaan tiap cluster maka harus dilakukan uji ANOVA. Bila menggunakan F hitung dalam analisis klaster, aturannya adalah semakin tinggi F hitung (dalam hal uji hipotesis, F hitung lebih besar dari F-tabel) dan tingkat signifikansi ($\text{sig} < 0,05$; semakin besar perbedaan antara keempat cluster yang terbentuk. Berikut hasil uji ANOVA.

Tabel 4.9 Hasil Uji ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore: X1	14.760	3	.793	199	18.624	.000
Zscore: X2	2.851	3	.972	199	2.932	.035
Zscore: X3	21.912	3	.685	199	32.001	.000
Zscore: X4	44.521	3	.344	199	129.459	.000
Zscore: X5	4.236	3	.951	199	4.453	.005
Zscore: X6	59.206	3	.123	199	483.203	.000
Zscore: X7	56.165	3	.168	199	333.587	.000
Zscore: X8	30.881	3	.550	199	56.194	.000
Zscore: X9	12.286	3	.830	199	14.805	.000
Zscore: X10	14.678	3	.794	199	18.491	.000
Zscore: X11	10.538	3	.856	199	12.308	.000
Zscore: X12	43.241	3	.363	199	119.057	.000
Zscore: X13	4.722	3	.944	199	5.002	.002

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Berdasarkan tabel 4.9, pada kolom cluster adalah *between cluster means* dan kolom error adalah *within cluster means*. Sementara pada kolom F didapat dari rumus sebagai berikut.

$$F = \frac{\text{between cluster means}}{\text{within cluster means}}$$

Hipotesis:

H_0 = Keempat *cluster* tidak mempunyai perbedaan yang signifikan

H_1 = Keempat *cluster* mempunyai perbedaan yang signifikan

Jika angka sig. > 0,05; H_0 diterima H_1 ditolak

Jika angka sig. < 0,05; H_0 ditolak H_1 diterima

Berdasarkan hasil yang diperoleh seluruh variabel mempunyai nilai signifikan (sig.) < 0,05 maka keempat cluster

mempunyai perbedaan yang signifikan dan hasilnya variabel X₆ jumlah SKS merupakan variabel yang paling membedakan anggota keempat cluster karena mempunyai nilai F tertinggi dari variabel-variabel tersebut.

1.2 Interpretasi K-Means Cluster

Pada keluaran akhir pusat kluster dapat dilihat informasi terkait proses standardisasi berdasarkan z-score.

**Tabel 4.10 Hasil Cluster
Final Cluster Centers**

	Cluster			
	1	2	3	4
Zscore: X1	-.05445	-.01007	-.55883	.61041
Zscore: X2	-.27343	.22952	-.11760	.20769
Zscore: X3	.98444	.21146	-.48310	-.29736
Zscore: X4	1.44374	.08764	-.70267	-.34119
Zscore: X5	.43692	-.22727	-.20071	-.00038
Zscore: X6	-1.74701	.54496	.39619	.57906
Zscore: X7	-1.62038	.86453	.08218	.66413
Zscore: X8	-.49127	.98864	-.68752	.59572
Zscore: X9	-.54088	-.18398	.57824	-.12361
Zscore: X10	-.01884	.55819	-.62199	.39203
Zscore: X11	-.01224	-.48185	.53375	-.31514
Zscore: X12	.10405	-1.92721	.30124	.49416
Zscore: X13	.24196	-.08966	-.35089	.22687

Tabel 4.10 menunjukkan hasil proses clustering akhir yang menghasilkan 4 cluster untuk setiap variabel. Variabel-variabel pada tabel pusat cluster akhir merupakan hasil dari nilai standarisasi. Angka negatif berarti data berada di bawah rata-rata umum dan angka positif berarti data berada di atas rata-rata umum (Desy, 2021). Berdasarkan Tabel 4.10 didapatkan karakteristik pada masing-masing cluster ialah:

- a. *Cluster 1* berisi variabel X₃, X₄, X₅, X₁₂, X₁₃, berada di atas rata-rata total serta X₁, X₂, X₆, X₇, X₈, X₉, X₁₀, X₁₁ berada di bawah rata-rata total.

- b. *Cluster 2* berisi variabel $X_2, X_3, X_4, X_6, X_7, X_8, X_{10}$ berada di atas rata-rata total serta $X_1, X_5, X_9, X_{11}, X_{12}, X_{13}$ berada di bawah rata-rata total.
- c. *Cluster 3* berisi variabel $X_6, X_7, X_9, X_{11}, X_{12}$, berada di atas rata-rata total serta $X_1, X_2, X_3, X_4, X_5, X_8, X_{10}, X_{13}$ berada di bawah rata-rata total.
- d. *Cluster 4* berisi variabel $X_1, X_2, X_6, X_7, X_8, X_{10}, X_{12}, X_{13}$ berada di atas rata-rata total serta $X_3, X_4, X_5, X_9, X_{11}$ berada di bawah rata-rata total.

Statistik deskriptif hasil cluster yang diperoleh pada Lampiran 4 disajikan pada Tabel 4.11 di bawah ini.

Tabel 4.11 Statistik Deskriptif Variabel

Variabel		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Kabupaten	Kabupaten Nias	7	7	13	2
	Kabupaten Nias Barat	4	1	15	3
	Kabupaten Nias Selatan	9	1	13	1
	Kabupaten Nias Utara	7	7	19	12
	Kota Gunungsitoli	18	13	5	45
	Yang Lain: Kabupaten Tapanuli Tengah	0	0	0	1
Asal Sekolah	SMA	33	15	45	38
	SMK	12	14	20	24
	MA	0	0	0	2
Usia	Mean	22,3	21,1	20,2	20,4
Semester	2	0	3	28	14
	4	0	7	28	30
	6	0	19	9	20
	8	45	0	0	0

Durasi Belajar	Kurang dari 1 jam	18	9	21	12
	1 - 2 jam	12	12	26	27
	2 - 3 jam	5	6	11	22
	3 - 4 jam	9	2	7	2
	> 4 Jam	1	0	0	1
Jumlah SKS	Mean	13	23,06	22,41	23,21
IPS	Mean	1,77	3,40	2,90	3,27
IPK	Mean	2,98	3,37	2,93	3,27
Memiliki Laptop	Ya	43	23	29	49
	Tidak	2	6	36	15
Tempat Tinggal	Kos	20	6	52	11
	Rumah	21	17	10	45
	Asrama	1	2	0	0
	Tempat Saudara	3	4	3	8
Memiliki Kendaraan	Ya	22	21	14	41
	Tidak	23	8	51	23
Beasiswa	Ya	7	28	5	0
	Tidak	38	1	60	64
Pendapatan Orangtua	Kurang dari Rp. 500.000	18	12	35	21
	Rp.500.000 – Rp. 1.000.000	12	11	25	22
	Rp. 1.000.000 – Rp. 2.000.000	5	5	3	12
	Diatas Rp. 2.000.000	10	1	2	9

Cluster 1 untuk variabel Kabupaten terdiri atas 7 orang Kabupaten Nias, 4 orang Kabupaten Nias Barat, 9 orang Kabupaten Nias Selatan, 7 orang Kabupaten Nias Utara, dan 18 orang di Kota Gunungsitoli. Variabel asal sekolah terdiri atas 33 orang SMA dan 12 orang SMK. Rata-rata usia mahasiswa di cluster 1 adalah 22 tahun dan hanya diisi oleh semester 8 sebanyak 45 orang. Dengan rata-rata durasi belajar 18 orang kurang dari

jam, 12³ orang 1-2 jam, 5 orang 2-3 jam, 9 orang 3-4 jam serta 1 orang lebih dari 4 jam. Jumlah rata-rata SKS adalah 13 dengan IPS rata-rata 1,77 dan IPK rata-rata 2,98. Yang memiliki laptop 43 orang dan yang tidak 2 orang. Tempat tinggal kos sebanyak 20 orang, rumah sebanyak 21 orang, asrama 1 orang serta 3 orang di tempat saudara. Yang memiliki kendaraan khususnya sepeda motor 22 orang dan yang tidak sebanyak 23 orang. Penerima beasiswa sebanyak 7 orang dan yang tidak sebanyak 38 orang. Pendapatan orangtuanya⁴ kurang dari Rp. 500.000 sebanyak 18 orang, kisaran Rp.500.000 – Rp. 1.000.000 sebanyak 12 orang, untuk kisaran Rp. 1.000.000 – Rp. 2.000.000 sebanyak 5 orang dan yang diatas Rp. 2.000.000 sebanyak 10 orang mahasiswa.

Cluster 2 untuk variabel Kabupaten terdiri atas 7¹⁰ orang Kabupaten Nias, 1 orang Kabupaten Nias Barat, 1 orang Kabupaten Nias Selatan, 7 orang Kabupaten Nias Utara, dan 13 orang di Kota Gunungsitoli. Variabel asal sekolah terdiri atas 15 orang SMA dan 14 orang SMK. Rata-rata usia mahasiswa di cluster 2 adalah 21 tahun dan diisi oleh semester 2 sebanyak 3 orang, semester 4 sebanyak 7 orang dan semester 6 sebanyak 19 orang. Dengan rata-rata durasi belajar 9 orang kurang dari jam, 12 orang 1-2 jam, 6 orang 2-3 jam, 2 orang 3-4 jam³. Jumlah rata-rata SKS adalah 23 dengan IPS rata-rata 3,40 dan IPK rata-rata 3,37. Yang memiliki laptop 23 orang dan yang tidak 6 orang. Tempat tinggal kos sebanyak 6 orang, rumah sebanyak 17 orang, asrama 2 orang serta 4 orang di tempat saudara. Yang memiliki kendaraan khususnya sepeda motor 21 orang dan yang tidak sebanyak 8 orang. Penerima beasiswa sebanyak 28 orang dan yang tidak sebanyak 1 orang⁴. Pendapatan orangtuanya kurang dari Rp. 500.000 sebanyak 12 orang, kisaran Rp.500.000 – Rp. 1.000.000 sebanyak 11 orang, untuk kisaran Rp. 1.000.000 – Rp. 2.000.000 sebanyak 5 orang dan yang diatas Rp. 2.000.000 sebanyak 1 orang mahasiswa.

Cluster 3 untuk variabel Kabupaten terdiri atas 13 orang Kabupaten Nias, 15 orang Kabupaten Nias Barat, 13¹⁰ orang Kabupaten Nias Selatan, 19 orang Kabupaten Nias Utara, dan 5 orang di Kota Gunungsitoli. Variabel asal sekolah terdiri atas 45 orang SMA dan 20 orang SMK. Rata-rata usia mahasiswa di cluster 3 adalah 20 tahun dan diisi oleh semester 2 sebanyak

28 orang, semester 4 sebanyak 28 orang dan semester 6 sebanyak 9 orang. Dengan rata-rata durasi belajar 21 orang kurang dari jam, 26 orang 1-2 jam, 11 orang 2-3 jam, 7 orang 3-4 jam. Jumlah rata-rata SKS adalah 22 dengan IPS rata-rata 2,90 dan IPK rata-rata 2,93. Yang memiliki laptop 29 orang dan yang tidak 36 orang. Tempat tinggal kos sebanyak 52 orang, rumah sebanyak 10 orang serta 3 orang di tempat saudara. Yang memiliki kendaraan khususnya sepeda motor 14 orang dan yang tidak sebanyak 51 orang. Penerima beasiswa sebanyak 5 orang dan yang tidak sebanyak 60 orang. Pendapatan orangtuanya kurang dari Rp. 500.000 sebanyak 35 orang, kisaran Rp.500.000 – Rp. 1.000.000 sebanyak 25 orang, untuk kisaran Rp. 1.000.000 – Rp. 2.000.000 sebanyak 3 orang dan yang diatas Rp. 2.000.000 sebanyak 2 orang mahasiswa.

Cluster 4 untuk variabel Kabupaten terdiri atas 2 orang Kabupaten Nias, 3 orang Kabupaten Nias Barat, 1 orang Kabupaten Nias Selatan, 12 orang Kabupaten Nias Utara, 45 orang di Kota Gunungsitoli dan 1 orang di luar yaitu Kabupaten Tapanuli Tengah. Variabel asal sekolah terdiri atas 38 orang SMA dan 24 orang SMK. Rata-rata usia mahasiswa di cluster 4 adalah 20 tahun dan diisi oleh semester 2 sebanyak 14 orang, semester 4 sebanyak 30 orang dan semester 6 sebanyak 20 orang. Dengan rata-rata durasi belajar 12 orang kurang dari jam, 27 orang 1-2 jam, 22 orang 2-3 jam, 2 orang 3-4 jam serta 1 orang yang lebih dari 4 jam. Jumlah rata-rata SKS adalah 23 dengan IPS rata-rata 3,27 dan IPK rata-rata 3,27. Yang memiliki laptop 49 orang dan yang tidak 15 orang. Tempat tinggal kos sebanyak 11 orang, rumah sebanyak 45 orang, serta 8 orang di tempat saudara. Yang memiliki kendaraan khususnya sepeda motor 41 orang dan yang tidak sebanyak 23 orang. Penerima beasiswa tidak ada serta yang tidak sebanyak 64 orang. Pendapatan orangtuanya kurang dari Rp. 500.000 sebanyak 21 orang, kisaran Rp.500.000 – Rp. 1.000.000 sebanyak 22 orang, untuk kisaran Rp. 1.000.000 – Rp. 2.000.000 sebanyak 12 orang dan yang diatas Rp. 2.000.000 sebanyak 9 orang mahasiswa.

BAB V SIMPULAN DAN SARAN

1.7 Simpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan, maka dapat disimpulkan bahwa:

- a. Apabila data mahasiswa program studi matematika Universitas Nias dikumpulkan dengan menerapkan algoritma *K-Means Clustering*, diperoleh 4 cluster, dimana mahasiswa dengan IPS yang tinggi masuk ke Cluster 2 dengan rata-rata 3.40. Mahasiswa dengan IPS yang sedang masuk ke Cluster 4 dengan rata-rata 3.27, Mahasiswa dengan IPS yang rendah masuk ke Cluster 3 dengan rata-rata 2.90, dan Mahasiswa dengan IPS yang sangat rendah masuk ke Cluster 4 dengan rata-rata 1.77.
- b. Mahasiswa semester 8 semuanya termasuk dalam cluster yang sama dibandingkan dengan mahasiswa semester lainnya. Hal ini disebabkan oleh pengaruh mata kuliah yang telah diselesaikan dan hasil IPS. Rata-rata nilai IPS pada semester 8 biasanya lebih rendah dibandingkan dengan nilai rata-rata mahasiswa pada semester lainnya.
- c. Keempat cluster yang terbentuk memiliki perbedaan signifikan dibuktikan dengan nilai uji ANOVA signifikan < 0.05 dan variabel yang paling berpengaruh dalam pengelompokan adalah jumlah SKS, IPS, semester, beasiswa, IPK, umur, wilayah, tempat tinggal, kepemilikan laptop, kepemilikan kendaraan, pendapatan orang tua, rata-rata pendidikan, dan terakhir sekolah asal.

1.8 Saran

Mengenai penelitian selanjutnya, sebaiknya dilakukan penelitian lebih lanjut dengan menggunakan metode clustering lain, seperti menggabungkan metode hierarki dan non-hierarki atau teknik multivariat lainnya. Atau anda akan mendapatkan hasil berbeda dengan metode berbeda yang digunakan dalam analisis cluster. Dan menggabungkan beberapa metode analisis massa

dengan metode statistik sangat diperlukan untuk mendapatkan solusi yang optimal.

ANALISIS CLUSTER MAHASISWA PENDIDIKAN MATEMATIKA UNIVERSITAS NIAS

ORIGINALITY REPORT

14%

SIMILARITY INDEX

15%

INTERNET SOURCES

10%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	lib.unnes.ac.id Internet Source	4%
2	prosiding.unimus.ac.id Internet Source	2%
3	ojs.unm.ac.id Internet Source	2%
4	repo.uinsatu.ac.id Internet Source	2%
5	Submitted to Institut Pertanian Bogor Student Paper	1%
6	id.scribd.com Internet Source	1%
7	journal.unnes.ac.id Internet Source	1%
8	repository.uinsu.ac.id Internet Source	1%
9	pt.scribd.com Internet Source	1%

10

bangimam-berbagi.blogspot.com

Internet Source

1%

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On