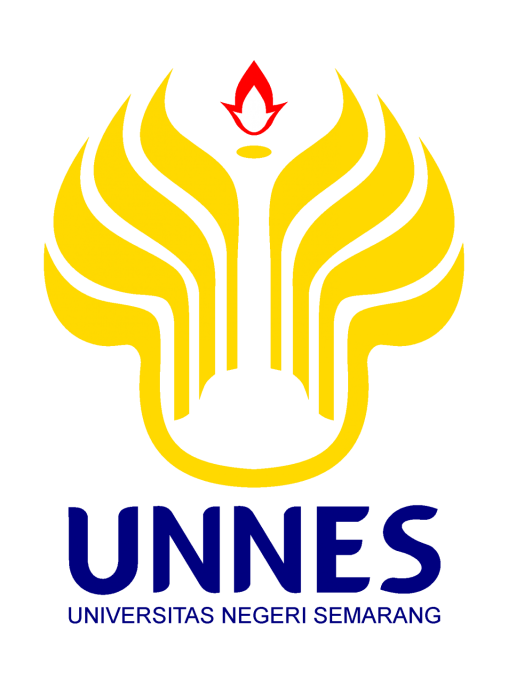
**MAKALAH STATISTIKA MULTIVARIAT**

**ANALISIS CLUSTER DENGAN METODE K-MEANS**

**(UNSUPERVISED LEARNING)**



**Di Susun Oleh :**

**Reza Dilla Saputri (4112317010)**

**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS NEGERI SEMARANG**

**2019**

# **Kata Pengantar**

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat-Nya sehingga makalah ini dapat tersusun hingga selesai .Tidak lupa kami juga mengucapkan banyak terimakasih atas bantuan dari pihak yang telah berkontribusi dengan memberikan sumbangan baik materi maupun pikirannya.

Dan harapan kami semoga makalah ini dapat menambah pengetahuan dan pengalaman bagi para pembaca, Untuk ke depannya dapat memperbaiki bentuk maupun menambah isi makalah agar menjadi lebih baik lagi.

Karena keterbatasan pengetahuan maupun pengalaman kami, Kami yakin masih banyak kekurangan dalam makalah ini, Oleh karena itu kami sangat mengharapkan saran dan kritik yang membangun dari pembaca demi kesempurnaan makalah ini.

Semarang, 7 Juni 2019

Penyusun

**DAFTAR ISI**

[**Kata Pengantar** 2](#_Toc12618381)

[BAB I 4](#_Toc12618382)

[PENDAHULUAN 4](#_Toc12618383)

[1.1. Latar Belakang 4](#_Toc12618384)

[1.2. Rumusan Masalah 5](#_Toc12618385)

[1.3. Tujuan 5](#_Toc12618386)

[BAB II 6](#_Toc12618387)

[PEMBAHASAN 6](#_Toc12618388)

[**1.** **Pengertian Analisis Cluster** 6](#_Toc12618389)

[**2.** **Ciri Ciri Cluster yang Baik** 6](#_Toc12618390)

[**3.** **Cara Kerja Analisis Cluster** 7](#_Toc12618391)

[**4.** **Proses Analisis Cluster** 7](#_Toc12618392)

[**5.** **Contoh Penerapan Analisis Cluster di R-Studio** 10](#_Toc12618393)

[BAB III 17](#_Toc12618394)

[PENUTUP 17](#_Toc12618395)

[Saran 17](#_Toc12618397)

[DAFTAR PUSTAKA 18](#_Toc12618398)

BAB I

PENDAHULUAN

* 1. Latar Belakang

Unsupervised learning merupakan pembelajaran yang tidak terawasi dimana tidak memerlukan target output. Pada metode ini tidak dapat ditentukan hasil seperti apa yang diharapkan selama proses pembelajaran, nilai bobot yang disusun dalam proses range tertentu tergantung pada nilai output yang diberikan. Tujuan metode uinsupervised learning ini agar kita dapat mengelompokkan unit-unit yang hampir sama dalam satu area tertentu. Pembelajaran ini biasanya sangat cocok untuk klasifikasi pola. Contoh algoritma jaringan saraf tiruan yang menggunakan metode unsupervised ini adalah competitive, hebbian, kohonen, LVQ (*Learning Vector Quantization)*, neocognitron.Salah satu contoh analisis menggunakan metode unsupervised learning adalah cluster analysis.

Clustering memberikan contoh yang mudah. Pertimbangkan kasus di mana input adalah kegiatan fotoreseptor yang dibuat oleh berbagai gambar apel atau jeruk. Dalam waktu semua kegiatan yang mungkin, ini masukan tertentu membentuk dua kelompok, dengan banyak sedikit de-grees variasi dibandingkan, yaitu dimensi yang lebih rendah. Salah satu tugas alami untuk belajar tanpa pengawasan adalah untuk menemukan dan mengkarakterisasi ini terpisah, cluster dimensi rendah.

Cluster atau kluster dapat diartikan ‘kelompok’ dengan demikian ,pada dasarnya analisis cluster pada dasarnya akan menghasilkan sejumlah kluster(kelompok).Analisis ini diawali dengan pemahaman bahwa sejumlah data tertentu sebenarnya mempunyai kemiripan diantara anggotanya .Karena itu,dimungkinkan untuk mengelompokan anggota anggota yang ‘mirip’ atau mempunyai karakteristik yang serupa tersebut dalam satu atau lebih kluster.

Analisis cluster bisa diterapkan pada banyak bidang ilmu,seperti psikologi (melakukan pengelompokan orang berdasar respons mereka terhadap stimulus tertentu,atau pengelompokan orang berdasar kepribadian mereka) ,biologi (membantu proses taksonomi untuk mengelompokan organisme tertentu) ,manajemen (membantu mengelompokan konsumen berdasar pendapat mereka terhadap produk tertentu) ,kedokteran (membantu menggolongkan sejumlah penyakit berdasar ciri ciri dari peyakit tersebut).Dan pada makalah ini akan dibahas tentang analisis cluster menggunakan metode K-Means

* 1. Rumusan Masalah

1. Apa yang dimaksud dengan analisis cluster?
2. Bagaimana analisis cluster yang baik dengan metode K-means?
3. Bagaimana cara kerja analisis cluster?
4. Bagaimana proses analisis cluster dengan motode K-Means?
5. Bagaimana contoh penerapan dari analisis cluster K-means di R Studio?
   1. Tujuan
6. Untuk mengetahui tentang apa itu analisis cluster
7. Untuk mengetahui ciri analisis cluster yang baik dengan metode K-means
8. Untuk mengetahui cara kerja analisis cluster
9. Untuk mengetahui proses analisis cluster
10. Untuk mengetahui contoh penerapan analisis cluster k-means di R

BAB II

PEMBAHASAN

1. **Pengertian Analisis Cluster**

Analisis cluster adalah suatu analisis statistik yang bertujuan memisahkan obyek kedalam beberapa kelompok yang mempunyai sifat berbeda antar kelompok yang satu dengan yang lain. Dalam analisis ini tiap-tiap kelompok bersifat homogen antar anggota dalam kelompok atau variasi obyek dalam kelompok yang terbentuk sekecil mungkin.

1. **Ciri Ciri Cluster yang Baik**

Ciri sebuah cluster yang baik

Seperti diketahui.analisis cluster akan membagi sejumlah data pada satu atau beberapa cluster tertentu.

Secara logika,sebuah cluster yang baik adalah cluster yang mempunyai

1.homogenitas (kesamaan) yang tinggi antar anggota dalam satu cluster (within-cluster).sebagai contoh,cluster konsumen rumah yang peduli lingkungan tentu terdiri dari orang orang yang mengutamakan kebersihan dan kenyamanan lingkungan rumahnya,mereka yang mengutamakan harga rumah murah tentu tidak dapat digabungkan menjadi ‘anggota’ cluter tersebut

2.homogenitas (perbedaan) yang tinggi antar cluster yang satu dengan cluster lainnya (between cluster).Dalam contoh diatas ,anggota dari cluster konsumen rumah yang peduli lingkungan tentu mempunyai pendapat yang jelas berbeda dengan anggota anggota cluster konsumen rumah yang mementingkan harga murah

Dari dua hal diatas,dapat disimpulkan bahwa sebuah cluster yang baik adalah cluster yang mempunyai anggota anggota yang semirip mungkin satu dengan yang lain,namun sangat tidak mirip dengan anggota anggota cluster yang lain.Di sini , ‘mirip’ diartikan sebagai tingkat kesamaan karakteristik antara dua data.

1. **Cara Kerja Analisis Cluster**

Secara garis besar ada tiga hal yang harus terjawab dalam proses kerjaanalisis cluster, yaitu :

1. Bagaimana mengukur kesamaan ?

Ada tiga ukuran untuk mengukur kesamaaan antar objek, yaitu ukurankorelasi, ukuran jarak, dan ukuran asosiasi.

1. Bagaimana membentuk cluster ?

Prosedur yang diterapkan harus dapat mengelompokkan objek-objek yangmemiliki kesamaan yang tinggi ke dalam sutau cluster yang sama.

1. Berapa banyak cluster/kelompok yang akan dibentuk ?

Pada prinsipnya jika jumlah cluster berkurang maka homogenitas alamcluster secara otomatis akan menurun.

1. **Proses Analisis Cluster**

Sebagaimana teknik multivariat lain proses analisis cluster dapatdijelaskan dalam enam tahap sebagai berikut :

4.1 Tahap Pertama : Tujuan Analisi Cluster

Tujuan utama analisis cluster adalah mempartisi suatu set objek menjadi dua kelompok atau lebih berdasarkan kesamaan karakteristik khusus yang dimilikinya.Dalam pembentukan kelompok/cluster dapat dicapai tiga tujuan, yaitu

1. Deskripsi klasifikasi (*taxonomy description*)

Penerapan anallisis cluster secara tradisisonal bertujuan mengeksplorasi dan membentuk suatu klasisfikasi/taksonomi secara empiris. Karena kemampuan partisinya analisis cluster dapatditerapkan secara luas. Meskipun secara empiris merupakan teknik eksplorasi analisis cluster dapat pula digunakan untuk tujuan konfirmasi.

4.2 Tahap Kedua : Desain Penelitian dalam Analisis Cluster

Tiga hal penting dalam tahap ini adalah pendeteksian outlier,mengukur kesamaan, dan standarisasi data.

1. Pendeteksian Outlier

Outlier adlah suatu objek yang sangat berbeda dengan objek lainnya.Outlier dapat digambarkan sebagai observasi yang secara nyata kebiasaan, tidak mewakili populasi umum, dan adanya under sampling dapat pula memunculkan outlier. Outlier menyebabkan struktur yang tidak benar dan cluster yang terbentuk menjadi tidak representatif.

1. Mengukur Kesamaan antar Objek

Konsep kesamaan adalah hal yang fundamental dalam analisis cluster.Kesamaan antar objek merupakan ukuran korespondensi antar objek.Ada tiga metode yang dapat diterapkan, yaitu ukuran korelasi, ukuran jarak, dan ukuran asosiasi.

1. Standarisasi Data
2. Standarisasi Variabel

Bentuk paling umum dalam standarisasi variabel adalah konversisetiap variabel terhadap skor atandar (dikenal dengan Z score) dengan melakukan substraksi nilai tengan dan membaginya dengans tandar deviasi tiap variabel.

1. Standarisasi Data

Berbeda dengan standarisasi variabel, standarisasi ndata dilakukanterhadap observasi/objek yang akan dikelompokkan.

4.3 Tahap Ketiga : Asumsi-asumsi dalam Analisis Cluster

Seperti hal teknik analisis lain,analisis cluster juga menetapkanadanya suatu asumsi. Ada dua asumsi dalam analisis cluster, yaitu :

1. Kecukupan Sampel untuk merepresentasikan/mewakili Populasi

Biasanya suatu penelitian dilakukan terhadap populasi diwakili oleh ekelompok sampel. Sampel yang digunakan dalam analisis cluster harus dapat mewakili populasi yang ingin dijelaskan, karena analisisini baik jika sampel representatif. Jumlah sampel yang diambiltergantung penelitinya, seorang peneliti harus yakin bahwa sampilyang diambil representatif terhadap populasi.

1. Pengaruh Multukolinieritas

Ada atau tidaknya multikolinieritas antar variabel sangatdiperhatikan dalam analisis cluster karena hal itu berpengaruh,sehingga variabel-variabel yang bersifat multikolinieritas secaraeksplisit dipertimbangkan dengan lebih seksama.

4.4 Tahap Keempat : Proses Mendapatkan Cluster dan Menilai kelayakan secara keseluruhan

Ada dua proses penting yaitu algoritma cluster dalam pembentukancluster dan menentukan jumlah cluster yang akan dibentuk. Keduanyamempunyai implikasi substansial tidak hanya pada hasil yang diperolehtetapi juga pada interpretasi yang akan dilakukan terhadap hasil tersebut.

Algoritma Cluster

Algoritma cluster harus dapat memaksimalkan perbedaan relatif cluster terhadap variasi dalam cluster. Dua metode paling umum dalamalgoritma cluster adalahmetode hirarkhi dan metode non hirarkhi.Penentuan metode mana yag akan dipakai tergantung kepada peneliti dankonteks penelitian dengan tidak mengabaikan substansi, teori dan konsepyang berlaku.

Keduanya memiliki kelebihan sendiri-sendiri. Keuntungan metode hirarkhi adalah cepat dalam proses pengolahan sehingga menghemat waktu, namun kelemahannya metode ini dapat menimbulkan kesalahan. Selain itu tidak baik diterapkan untuk menganalisis sampeldengan ukuran besar. Metode Non Hirarkhi memiliki keuntungan lebihdaripada metode hirarkhi. Hasilnya memiliki sedikit kelemahan pada data outlier, ukuran jarak yang digunakan, dan termasuk variabel tak relevanatau variabel yang tidak tepat. Keuntungannya hanya dengan menggunakan titik bakal nonrandom, penggunaan metode non hirarkhi untuk titik bakal random secara nyata lebih buruk dari pada metodehirarkhi.

Alternatif lain adalah dengan mengkombinasikan kedua metode ini. Pertama gunakan metode hirarkhi kemudian dilanjutkan dengan metode non hirarkhi

1. Metode Hirarkhi

Tipe dasar dalam metode ini adalah aglomerasi dan pemecahan. Dalam metode aglomerasi tiap observasi pada mulanya dianggap sebagai cluster tersendiri sehingga terdapat cluster sebanyak jumlah observasi. Kemudian dua cluster yang terdekat kesamaannyadigabung menjadi suatu cluster baru, sehingga jumlah cluster berkurang satu pada tiap tahap. Sebaliknya pada metode pemecahan dimulai dari satu cluster besar yang mengandung seluruh observasi,selanjutnya observasi-observasi yang paling tidak sama dipisah dan dibentuk cluster-cluster yang lebih kecil. Proses ini dilakukan hinggatiap observasi menjadi cluster sendiri-sendiri.Hal penting dalam metode hirarkhi adalah bahwa hasil padatahap sebelumnya selalu bersarang di dalam hasil pada tahap berikutnya, membentuk sebuah pohon.

1. Metode Non Hirarkhi

Berbeda dengan metode hirarki,metode ini justru dimulai dengan menentukan terlebih dahulu jumlh cluster yang diinginkan (dua cluster,tiga cluster,atau yang lain).Setelah jumlah cluster diketahui,baru proses cluster dilakukan tanpa mengikuti proses hirarki.Metode ini biasa disebut dengan K-Means Cluster.

4.5 Tahap Kelima : Interpretasi terhadap Cluster

Tahap interpretasi meliputi pengujian tiap cluster dalam term untuk menamai dan menandai dengan suatu label yang secara akurat dapat menjelaskan kealamian cluster. Proes ini dimulai dengan suatu ukuran yang sering digunakan yaitu centroid cluster. Membuat profil dan interpretasi cluster tidak hanya tidak hanya untuk memperoleh suatu gambaran saja melainkan pertama, menyediakan suatu rata-rata untuk menilai korespondensi pada cluster yang terbentuk,kedua, profil cluster memberikan araha bagi penilainan terhadap signifikansi praktis.

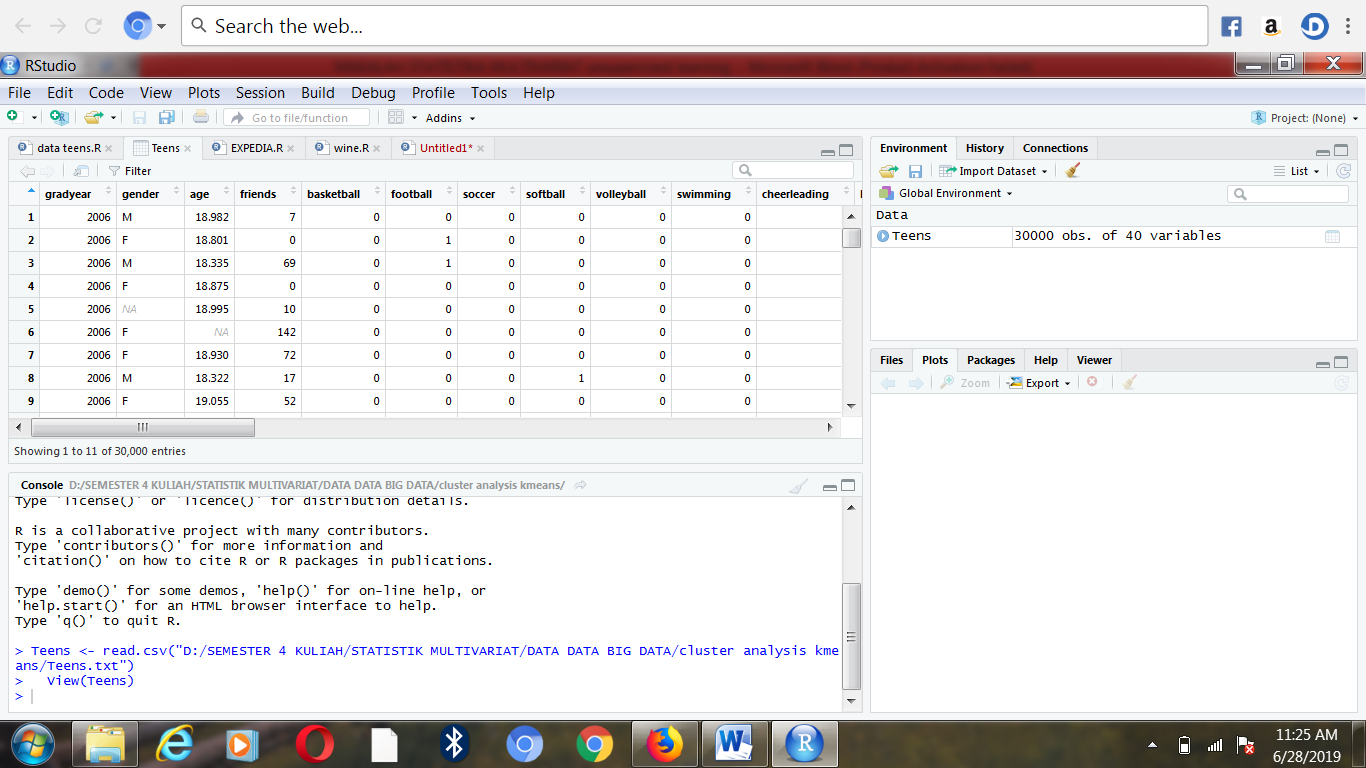
1. **Contoh Penerapan Analisis Cluster di R-Studio**

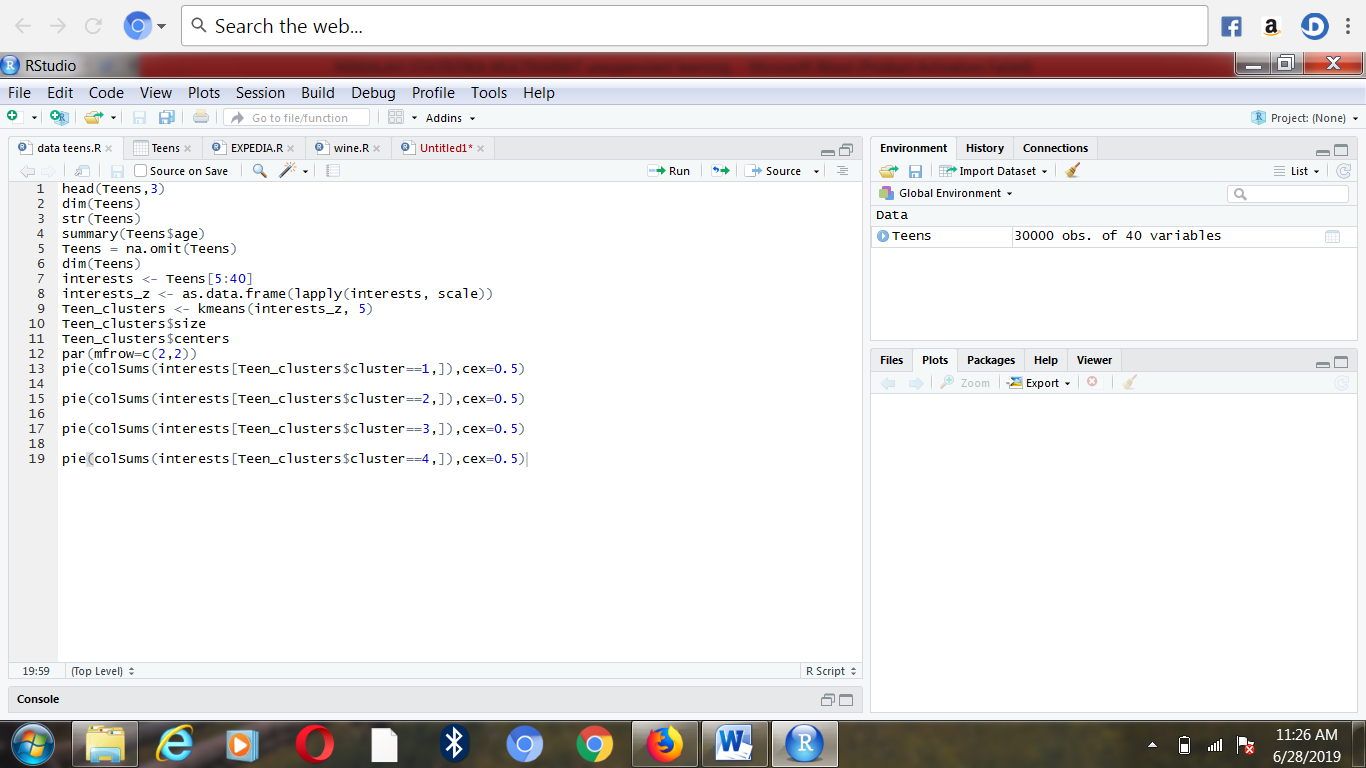
Untuk analisis ini, kami akan menggunakan dataset yang mewakili sampel acak 30.000 siswa sekolah menengah A.S. yang memiliki profil di Jejaring Sosial terkenal pada 2006 hingga 2009.Dari 500 kata teratas yang muncul di semua halaman, 36 kata dipilih untuk mewakili lima kategori minat, yaitu kegiatan ekstrakurikuler, mode, agama, romansa, dan perilaku antisosial. 36 kata termasuk istilah-istilah seperti sepak bola, seksi, ciuman, Alkitab, belanja, kematian, dan narkoba. Dataset terakhir menunjukkan, untuk setiap orang, berapa kali setiap kata muncul di profil SNS orang tersebut.

Isi dataset profil jejaring social pada tahun 2006 – 2009 :

* gradyear (tahun),gender(jenis kelamin),age(umur),friends(teman)
* lima kategori minat :

1. Ekstrakulikuler : basketball (bola basket) ,football (sepak bola), soccer, softball, volleyball, swimming, cheerleading,baseball,tennis,sports
2. Mode : cute,hair,dress,mall,shopping,blonde,clothes
3. Romansa : dance,band,marching,music,rock
4. agama : god,church,jesus,bible
5. perilaku antisosial : sex,sexy,kissed,drugs,hoilister,abercromble,die,deat,drunk

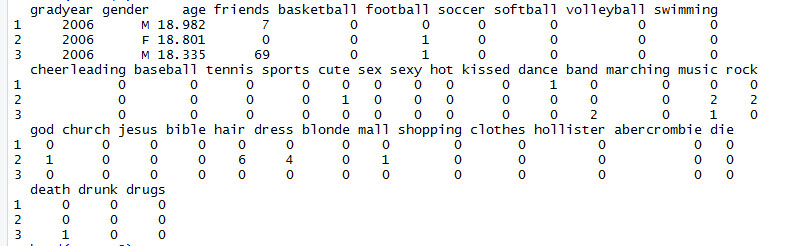




> Teens <- read.csv("D:/SEMESTER 4 KULIAH/STATISTIK MULTIVARIAT/DATA DATA BIG DATA/cluster analysis kmeans/Teens.txt")

> View(Teens)

> head(teens,3)



> dim(Teens)

[1] 30000 40

Mari kita juga melihat spesifik data. Beberapa baris pertama dari output str () adalah sebagai berikut :

> str(Teens)

'data.frame': 30000 obs. of 40 variables:

$ gradyear : int 2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...

$ gender : Factor w/ 2 levels "F","M": 2 1 2 1 NA 1 1 2 1 1 ...

$ age : num 19 18.8 18.3 18.9 19 ...

$ friends : int 7 0 69 0 10 142 72 17 52 39 ...

$ basketball : int 0 0 0 0 0 0 0 0 0 0 ...

$ football : int 0 1 1 0 0 0 0 0 0 0 ...

$ soccer : int 0 0 0 0 0 0 0 0 0 0 ...

$ softball : int 0 0 0 0 0 0 0 1 0 0 ...

$ volleyball : int 0 0 0 0 0 0 0 0 0 0 ...

$ swimming : int 0 0 0 0 0 0 0 0 0 0 ...

$ cheerleading: int 0 0 0 0 0 0 0 0 0 0 ...

$ baseball : int 0 0 0 0 0 0 0 0 0 0 ...

$ tennis : int 0 0 0 0 0 0 0 0 0 0 ...

$ sports : int 0 0 0 0 0 0 0 0 0 0 ...

$ cute : int 0 1 0 1 0 0 0 0 0 1 ...

$ sex : int 0 0 0 0 1 1 0 2 0 0 ...

$ sexy : int 0 0 0 0 0 0 0 1 0 0 ...

$ hot : int 0 0 0 0 0 0 0 0 0 1 ...

$ kissed : int 0 0 0 0 5 0 0 0 0 0 ...

$ dance : int 1 0 0 0 1 0 0 0 0 0 ...

$ band : int 0 0 2 0 1 0 1 0 0 0 ...

$ marching : int 0 0 0 0 0 1 1 0 0 0 ...

$ music : int 0 2 1 0 3 2 0 1 0 1 ...

$ rock : int 0 2 0 1 0 0 0 1 0 1 ...

$ god : int 0 1 0 0 1 0 0 0 0 6 ...

$ church : int 0 0 0 0 0 0 0 0 0 0 ...

$ jesus : int 0 0 0 0 0 0 0 0 0 2 ...

$ bible : int 0 0 0 0 0 0 0 0 0 0 ...

$ hair : int 0 6 0 0 1 0 0 0 0 1 ...

$ dress : int 0 4 0 0 0 1 0 0 0 0 ...

$ blonde : int 0 0 0 0 0 0 0 0 0 0 ...

$ mall : int 0 1 0 0 0 0 2 0 0 0 ...

$ shopping : int 0 0 0 0 2 1 0 0 0 1 ...

$ clothes : int 0 0 0 0 0 0 0 0 0 0 ...

$ hollister : int 0 0 0 0 0 0 2 0 0 0 ...

$ abercrombie : int 0 0 0 0 0 0 0 0 0 0 ...

$ die : int 0 0 0 0 0 0 0 0 0 0 ...

$ death : int 0 0 1 0 0 0 0 0 0 0 ...

$ drunk : int 0 0 0 0 1 1 0 0 0 0 ...

$ drugs : int 0 0 0 0 1 0 0 0 0 0 ...

Seperti yang kami harapkan, data tersebut mencakup 30.000 remaja dengan empat variabel yang menunjukkan karakteristik pribadi dan 36 kata yang menunjukkan minat. Perhatikan bahwa ada beberapa NA dalam variabel jenis kelamin.

> summary(Teens$age)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

3.086 16.312 17.287 17.994 18.259 106.927 5086

Kami akan melewati semua data dengan nilai yang hilang :

> Teens = na.omit(Teens)

> dim(Teens)

[1] 24005 40

Kami akan memulai analisis kluster kami dengan hanya mempertimbangkan 36 fitur yang mewakili berapa kali berbagai minat muncul di profil SNS remaja. Untuk kenyamanan, mari kita membuat bingkai data yang hanya berisi fitur-fitur ini :

> interests <- Teens[5:40]

Untuk menerapkan standardisasi z-skor pada kerangka data minat, kita dapat menggunakan fungsi scale () dengan lapply (), sebagai berikut :

> interests\_z <- as.data.frame(lapply(interests, scale))

Untuk membagi remaja menjadi lima kelompok, kita dapat menggunakan perintah berikut:

> Teen\_clusters <- kmeans(interests\_z, 5)

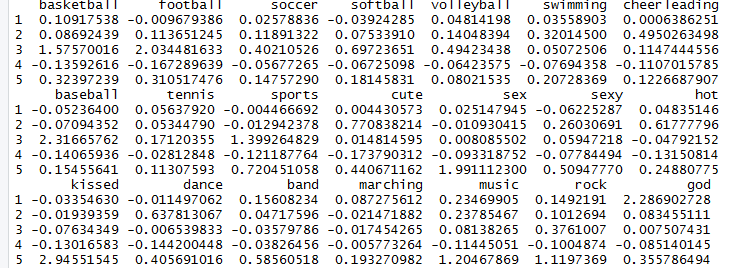
sejumlah contoh jatuh di masing-masing kelompok. Jika kelompok-kelompok itu terlalu besar atau terlalu kecil, maka mereka sepertinya tidak akan sangat berguna. Untuk mendapatkan ukuran kluster kmeans (), gunakan komponen $ teen\_clusters sebagai berikut: sejumlah contoh jatuh di masing-masing kelompok. Jika kelompok-kelompok itu terlalu besar atau terlalu kecil, maka mereka sepertinya tidak akan sangat berguna. Untuk mendapatkan ukuran kluster kmeans (), gunakan komponen $ teen\_clusters sebagai berikut:

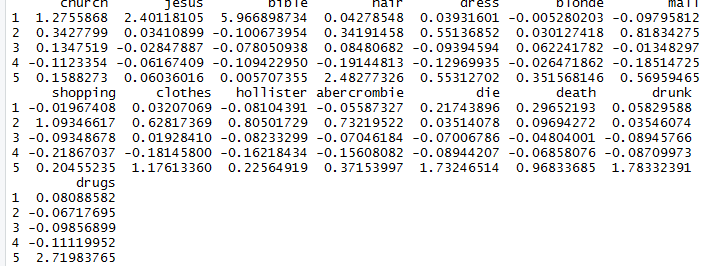
> Teen\_clusters$size

[1] 405 3553 1156 18036 855

Untuk melihat lebih dalam pada cluster, kita dapat memeriksa koordinat centroid cluster menggunakan komponen $ Pusat teen\_clusters, yang adalah sebagai berikut untuk delapan fitur pertama :

> Teen\_clusters$centers



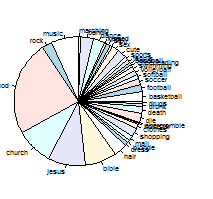


Output diatas adalah output dari clustering k means yang dibagi menjadi 5 cluster dimana cluster 1 terdiri dari basketball dengan rentang nilai 0,10917538,football dengan rentang nilai -0.009679386,soccer dengan rentang nilai 0,02578836,softball dengan rentang nilai -0,03924285,volleyball dengan rentang nilai 0,04814198,swimming dengan rentang nilai 0,03558903,cheerleading dengan rentang nilai 0,0006386386251,baseball dengan rentang nilai -0,05236400,tennis dengan rentang nnilai 0.05637920,sports dengan rentang nilai -0,004466692,cute dengan rentang nilai 0,004430573,sex dengan rentang nilai 0,025147945,sexy dengan rentang nilai -0,06225287,hot dengan rentang nilai 0,04835146,kissed dengan rentang nilai -0,03354630,dance dengan rentang nilai-0,011497062,band dengan rentang nilai 0,15608234,marching dengan rentang nilai 0,087275612,music dengan rentang nilai 0,1492191,rock dengan rentang nilai 0,1492191,god dengan rentang nilai 2,286902728,cnurcn dengan rentang nilai 1,2755868,jesus dengan rentang nilai 2,40118105,dioie dengan rentang nilai 5,966898734,hair dengan rentang nilai 0,04278548,dress dengan rentang nilai 0,03931601,dionae dengan rentang nilai -0,005280203,mall dengan rentang nilai -0,09795812 .Rentang tersebut adalah rentang untuk claster 1 begitu seterusnya sampai cluster 5.

Karakterisasi klaster dapat diperoleh dengan diagram lingkaran :

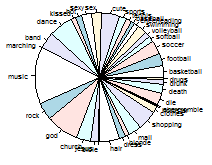
> par(mfrow=c(2,2))

> pie(colSums(interests[Teen\_clusters$cluster==1,]),cex=0.5)



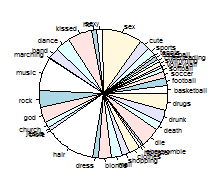
pada gambar pie chart diatas dapat dilihat bahwa profil di jejaring social yang paling banyak muncul di semua halaman adalah god (tuhan) yang artinya bahwa setiap orang yang yang akan masuk di jejaring social hal yang dicari adalah agama

> pie(colSums(interests[Teen\_clusters$cluster==2,]),cex=0.5)



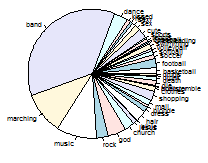
Pada gambar pie chart diatas dapat dilihat bahwa profil di jejaring social yang muncul di laman halaman pada alamat pencarian adalah music yang artinya bahwa kelompok kedua yang paling banyak dicari di jejaring social adalah music.

> pie(colSums(interests[Teen\_clusters$cluster==3,]),cex=0.5)



Pada gambar pie chart diatas dapat dilihat bahwa profil di jejaring social yang muncul di laman halaman pada alamat pencarian adalah model gaya rambut artinya bahwa kelompok 3 yang paling banyak dicari disitus pencarian adalah kelompok orang yang menyukai mode gaya rambut.

> pie(colSums(interests[Teen\_clusters$cluster==4,]),cex=0.5)



Pada gambar pie chart diatas dapat dilihat bahwa profil di jejaring social yang muncul di laman halaman pada alamat pencarian adalah band yang artinya bahwa kelompok selanjutnya yang paling banyak dicari disitu pecarian jejaring social adalah kelompok orang yang menyukai band.

BAB III

PENUTUP

Kesimpulan

Teori cluster analysis memiliki berbagai macam metode dengan cakupan yang luas sehingga terdapat banyak aplikasi dan implementasinya pada kehidupan sehari-hari. Analisis cluster dilakukan untuk mengelompokan objek-objek yang memiliki kemiripan (homogen). Berdasarkan karakteristik yangdimiliki,dengan analisis cluster sekelompok objek dapat dikelompokkan. Metode pengelompokan pada dasarnya ada dua, yaitu pengelompokan hirarki (Hierarchical Clustering Method) dan pengelompokan non hirarki (Non Hierarchical Clustering Method.Pengimplementasian formula pada contoh permasalahan pun tidak terlalu sulit, hanya saja memerlukan ketelitian dalam perhitungan. Cluster analysis juga memiliki kelebihan dan kekurangan. Keuntungan dari penggunaan metode ini adalah metode ini baik untuk meninjau pendataan dengan cepat, terutama jika benda tersebut diklasifikasikan ke dalam banyak kelompok. Namun kelemahannya adalah pada metode k-means diperlukan beberapa analisis sebelum jumlah cluster dapat ditentukan. Hal ini dapat sangat sensitive terhadap pilihan pusat awal cluster.

Saran

Diharapkan setelah membaca makalah ini para pembaca dapat mengetehui salah satu materi tentang analisis cluster dengan metode K-means yang merupakan salah satu dari teknik statistika multivariat menggunakan R-Studio. Menyadari bahwa makalah belum dari kata sempurna, kedepannya kami akan lebih fokus dan detail dalam menjelaskan tentang makalah di atas dengan sumber-sumber yang lebih banyak, yang tentunya dapat di pertanggungjawabkan.

DAFTAR PUSTAKA

Baumer,Kaplan,Horton\_Modern Data Science with R

<https://www.academia.edu/35587332/MAKALAH_ANALISA_CLUSTERING>

<https://rpubs.com/gabrielmartos/ClusterAnalysis>

<http://miftakhurrizal.lecture.ub.ac.id/files/2018/02/ANALISIS-CLUSTER.pdf>