**METODE *ENSEMBLE ROBUST CLUSTERING USING LINKS* (ROCK)**

**UNTUK PENGELOMPOKKAN PERGURUAN TINGGI SWASTA (PTS)**

**DI KOTA SEMARANG**

**Berliana Jannah1, Iut Tri Utami2, Arief Rachman Hakim3**

Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

**berlianajannah@gmail.com**

**DOI: 10.14710/J.GAUSS.XX.X.XX-XX**

|  |  |
| --- | --- |
| **Article Info:**Received: Accepted: Available Online: **Keywords:** *ROCK ensemble, PTS quality, mixed data* | **Abstract:** *The purpose of this research is to group PTS that have performance achievements in five years, through the quality of Human Resources and Students (Input), the quality of Institutional Management (process), the quality of Short-Term Performance Achievements (Output) and the quality of Long-Term Performance Achievements (Outcome). In addition, it can also be seen from the form of PTS, PTS Accreditation and PTS Research Performance. This PTS grouping uses mixed data, namely numerical data and categorical data. The method used for grouping mixed data is the ROCKeensemble methode(RObusteClusteringeusingelinKs). The results of clustering numerical data obtained the optimum number of groups 3, on categorical data obtained the optimum group 4. After clustering each type of data and merging and clustering obtained the optimum group 3 with a threshold (θ) is 0.2. The results of each group is: low quality consist of 29 PTS, medium quality consist of 7 PTS, and high quality there is 1 PTS. The results of this research can be used to cluster private universities in Semarang City, so that it can be used as a reference for prospective students in choosing private universities in Semarang, and can be referenced to the Central Java LLDIKTI in determining the quality of private universities in Semarang City.* |

1. **PENDAHULUAN**

Menurut UU Nomor 12 Tahun 2012, pendidikan tinggi di Indonesia dikategorikan 2 jenis, diantaranya (1) PerguruaneTinggieNegeri (PTN) dan (2) Perguruan Tinggi Swasta (PTS). Ketatnya seleksi penerimaan mahasiswa di PTN mengakibatkan tidak semua calon mahasiswa tertampung di PTN. Calon mahasiswa harus mendaftar PTS sebagai salah satu alternatif melanjutkan perguruan tinggi. Dalam memilih perguruan tinggi, calon mahasiswa mempertimbangkan pilihan lembaga PTS yang dituju untuk mendapatkan pendidikan yang memberikan kemudahan dalam mencari pekerjaan setelah lulus. Untuk mendapatkan perguruan tinggi yang memiliki prospek kedepan maka calon mahasiswa mempertimbangkan beberapa aspek, diantaranya (1) mutu SDM dan mahasiswa (*input*), (2) pengelolaanekelembagaan (proses), (3) pencapaian kinerja jangka pendek (*output*), dan (4) kinerja jangka panjang (*outcome*). Di samping itu dalam pemilihan PTS, calon mahasiswa juga mempertimbangkan bentuk PTS, akreditasi PTS, dan kinerja penelitian yang dilakukan oleh PTS.

Penelitian ini dimaksudkan untuk memberikan gambaran *cluster* perguruan tinggi di Kota Semarang, sebagai panduan mahasiswa dalam memilih PTS. Dengan mengelompokkan PTS di Kota Semarang yang bertujuan untuk mengetahui ciri khas PTS dan kualitas PTS yang berbeda. Metode untuk mengetahui kelompok PTS berdasarkan ciri khas dan kualitas menggunakan data campuran dengan metode *clustereensemble*, yaitu teknik pengelompokkan yang menggabungkan hasil dari pengelompokkan beberapa algoritma. Menurut He, et al (2005) dengan metode ini akan mendapatkan kelompok yang lebih baik.

1. **TINJAUANIPUSTAKA**
	1. **StatistikaiDeskriptif**

Statistikaedeskriptif adalah metode statistik yang berhubungan dengan pengumpulan data serta penyajian suatu informasi untuk membagikan data yang bermanfaat. Ada beberapa bentuk penyajian informasi diantaranya berupa grafik, gambar dan tabel. Penyajian informasi dapat menggunakan (1) rata-rata dan median dalam menjelaskan karakteristik data numerik. (2) bentuk grafik yang dapat menunjukkan frekuensierelatif data kategorik.

* 1. **AnalisisaCluster**

Analisis *cluster* merupakan analisiseemultivariat yang bertujuan untuk pengelompokkan *n* pengamatan ke dalam *C* kelompoke(*C ≤ n*) berdasar ciri khasnya. Hasil dari analisis ini dapat menjadi panduan dalam mengelompokkan data yang baru di *cluster* yang sebelumnya terbentuk. Menurut Johnson dan Winchern (2007) analisisa*cluster* merupakan metode primitif yang tidak dibutuhkan adanya asumsi yang dapat digunakan dalam mengelompokkan data dengan alasan pengelompokkan data berdasarkan kemiripan serta ketidakmiripan.

* + 1. **MetodeeHirarkie*Agglomerative***

Pengertian dari metode dalam pengelompokkan melalui dataenumerik dapat berdasarkan ukuran ketidakmiripan ataupun jarak. Sedangkan untuk ukuran ketidakmiripan yang sering digunakan ialah jarakie*uclidean,* yaituantar 2 pengamatan dan seberapa pengamatan sejumlah *n.* Berikutadalahrumus jarak *euclidean.*

$d\_{ij}=\sqrt{\left(x\_{i}-x\_{j}\right)'\left(x\_{i}-x\_{j}\right)}$ , $i, j=1,2,…, n dan i\ne j $ (1)

dengan $x\_{i}^{'}=[x\_{1i}, x\_{21}, x\_{3i}, …, x\_{mi}$], $x\_{j}^{'}=[x\_{1j}, x\_{2j}, x\_{3j}, …, x\_{mj}$].

Pengertian metode hirarki menurut Johnson dan Winchern (2007) merupakan metode pada objek yang mempunyai kesamaan ukuran yang tergabung dalam kelompok. Penggunaan metodeihirarki dilakukan apabila banyak kelompok yang sebelumnya tidak diketahui, serta pengamatan yang tidak terlalu besar. Di bawah ini merupakan metodeipengelompokkanihirarkii*agglomerative:*

1. *SingleeLinkage*

*Singleilinkage* sebagai pengelompokkan berdasarkan jarak terdekat dan kesamaan yang banyak. Rumus yang digunakan untuk menghitung ukuran jarak yang dapat digunakan antara kelompok (*UV*) dan *W* dapat ditunjukkan persamaan berikut:

$d\_{\left(UV\right)W }=min\{d\_{UW}, d\_{VW}\}$ (2)

1. *Complete Linkage*

*Completeilinkage* dapat diartikan sebagai metode *cluster* yang dibentuk melalui cara pengelompokkan objek-objek yang mempunyai jaraketerjauh atau yang memiliki kesamaan paling sedikit. Untuk mengetahui rumus ukuran jarak yang dipakai antara kelompok ke-*(UV)* dan ke *W* ditunjukkan pada persamaan (3).

$d\_{\left(UV\right)W }=maks\{d\_{UW}, d\_{VW}\}$ (3)

1. *AverageeLinkage*

*Average Linkage* didefinisikan sebagai metode cluster yang terbentuk melalui nilaierata-rata antar jarak beberapa individu ke dalam kelompok melalui rata-rata jarak dari sejumlah individu ke sejumlah individu lain. Peenggunaan rumus jarak antar kelompokeke-(UV) ke-W dapat menggunakan rumus sebagai berikut:

$$d\_{\left(UV\right)W}=\frac{1}{N\_{UV}N\_{W}}\sum\_{q}^{}\sum\_{r}^{}d\_{qr} (4)$$

* + 1. **Validasii*clusteri*MetodeiHirarki**

Jumlah dari kelompok optimum sesudah pengelompokkan dapat diketahui melalui validasie*cluster* di metode hirarki. Metode ini dapat dihitung melalui indekseR-*square*. Untuk mengetahui nilai *SST* ditunjukkan pada persamaan (5).

$$SST=\sum\_{i=l}^{m\_{numerik}}\sum\_{i=l}^{n}\left(x\_{il}-\overbar{x}\_{l}\right)^{2} \left(5\right)$$

Nilai *SumiSquareiWithiniGroupi(SSW)* dihitung dahulu sebelum menghitung nilai *SSB* dengan rumus ditunjukkan pada persamaaan (6).

$$SSW=\sum\_{c=l}^{C}\sum\_{i=l}^{m\_{numerik}}\sum\_{i=l}^{n}\left(x\_{ilc}-\overbar{x}\_{lc}\right)^{2} \left(6\right)$$

Nilai SSB dirumuskan dengan persamaan (7).

 $SSB=SST-SSW (7)$

Rumus Nilai *R-square* ditunjukkan pada persamaan (8)

$$R\_{square}= \frac{SSB}{SST}=\frac{[SST-SSW]}{SST} (2.8)$$

*R-square* dapat diartikan sebagai ukuran pembeda antarekelompok dengan sekitaran dari angka 0 sampaie1. *R-square* yang ditunjukkan dengan nilai 0 (nol) artinya tidak adanya perbedaan antar kelompok. Sedangkan untuk *R-square* yang ditunjukkan nilai 1 diartikan ada perbedaan besar antar kelompok.

* 1. **MetodeiROCK**

Penggunaan metode tradisional dengan konsep jarak antar titik dalam mengelompokkan variabel dengan tipe numerik menurut Guha et al (2000) dianggap tidak tepat apabila diterapkan dalam data kategorik. Maka dari itu, diciptakan metode ROCK dimana pengelompokkan dengan hirarki agglomerative yang bisa digunakan data kategorik.

Pembentukan konsep baru pada metode ROCK yaitu *link* yang digunakan untuk menghitung kedekatan atau kesamaan antar 2 titik data. Pengamatan yang memiliki hubungan rendah dipisahkan dengan data pengelompokkan, sedangkan untuk pengamatan yang memiliki hubungan yang tinggi dijadikan 1 kelompok. Syarat-syarat pengelompokkan algoritma ROCK akan berhenti ketika dalam keadaan (1) Total kelompok yang diinginkan terpenuhi, dan (2) Tidak adanya *link* antarikelompok.

Menurut Dutta et al (2005) langkah pengelompokkan dengan metode ROCK adalah dengan awal menentukan inisialisasi dengan data poin sebagai *cluster*, langkah selanjutnya dengan menghitungesimilaritas atau kesamaan kelompok.

Ukuran kemiripan antar pasanganepengamatan ke-*i* dan ke-*j* ditentukan dengan persamaan berikut:

$$sim\left(X\_{i}, X\_{j}\right)=\frac{\left|X\_{i}∩X\_{j}\right|}{\left|X\_{i}∩X\_{j}\right|} , i\ne j (2.9)$$

*i =* 1,2,3,….,n dan *j =* 1,2,3,…,n

*Xi*: Himpunanepengamatan ke-*i,* dengan $X\_{i}=\left\{x\_{1i},x\_{2i}, x\_{3i},…,x\_{m\_{kategori}i} \right\}$

*Xj*: Himpunanepengamatan ke-*j,* dengan $X\_{j}=\left\{x\_{1j},x\_{2j}, x\_{3j},…,x\_{m\_{kategori}j} \right\}$

1. Menentukan *neighbors*

Dalam pengamatan *Xi* yang dinyatakan *Xj* sebagai *neighbors* apabila nilai $sim\left(X\_{i}, X\_{j}\right)\geq θ$. *Threshold* *(θ)* merupakan parameter yang dapat ditentukan oleh peneliti sehingga bisa berfungsi untuk mengontrol kedekatan hubungan antar objek yang bernilai antara 0 sampai 1.

1. Menentukan *Link*

Untuk menentukan *Link* (*Xi*, *Xj*) antar objek bisa dihitung dari jumlahe*common neighbors* antara *Xi* dan *Xj* dengan ketentuan jika nilai *Xi* dan *Xj* besar maka kemungkinan *Xi* dan *Xj* ada pada *cluster* yang sama.

1. *GoodnesseMeasure*

Persamaan yang digunakan untuk menghitung total *link* dibagi dengan adanya kemungkinan *link* akan terbentuk berdasar ukuran kelompoknya (*Goodnessemeasure)*. Rumus persamaan dapat diketahui sebagai berikut:

$$g\left(C\_{i}, C\_{j}\right)=\frac{link\left(C\_{i}, C\_{j}\right)}{\left(n\_{i}+n\_{j}\right)^{1+2f\left(θ\right)}-n\_{i}^{1+2f\left(θ\right)}- n\_{j}^{1+2f\left(θ\right)}} (10)$$

$link\left(C\_{i}, C\_{j}\right)=\sum\_{X\_{i},ϵC, X\_{j},ϵC}^{}link(X\_{i}, X\_{j})$ merupakan jumlah *link* dari seluruh kemungkinan pasangan objek pada *Ci*dan *Cj.* Sedangkanni dan nj menunjukkan total anggota kelompok ke-*i* dan kelompok ke-*j* untuk $f\left(θ\right)=\frac{1-θ}{1+θ}$, dengan θ sebagai nilai *threshold* yang dipakai.

* 1. ***ClusteriEnsemble***

Metode *cluster* *ensemble*emerupakan metode penggabungan beberapa hasil yang didapatkan dari algoritmaepengelompokkan yangeberbeda. Tujuan dari metode ini untuk diperoleh solusi gabungan sebagai solusi akhir. Penggunaan pengelompokkan dari data numerik dan data kategorik dilakukan dengan cara membagiedata menurut jenis data secara terpisah. Dari hasil yang dikelompokkan selanjutnya digabungkan menggunakan Menurut He et al (2005), hasil dari algoritma pengelompokkan merupakan data kategorik yang selanjutnya dianggap sebagai pengelompokkan data kategorik.

Terdapat dua tahap algoritma pada pengelompokkan *ensemble*, yaitu (1) mengelompokkan sejumlah algoritma dan menyimpan hasil dari pengelompokkan, (2) dengan cara menentukan *final cluster* yang berasal dari kelompok yang sudah diperoleh.

* 1. **KinerjaeHasilePengelompokkan**

Kinerja hasil pengelompokkan bertujuan untuk mengetahui validitas kelompok. Menurut Hair et al (2001) kategori kelompok baik, harus memiliki keheterogenan yang sangat tinggi antar kelompok dan kehomogenan yang sangat tinggi dalam kelompok. Pengukuran kinerja yang dihasilkan data numerik dapat dihitung melalui rasio *Sw* dan *SB*. Nilai simpanganmbaku atau *within* (*SW*) menurut Bunker dan James (1996) ditunjukkan pada persamaan (11) berikut:

$$S\_{w}=\frac{1}{C}\sum\_{c=l}^{C}S\_{c} (11)$$

$$S\_{B}=\left[\frac{1}{C-1}\sum\_{c=l}^{C}(\overbar{x}\_{c}-\overbar{x})^{2}\right]^{1/2} (12) $$

Aturan yang digunakan dalam kinerja metode pengelompokkan dari data numerik sebagai berikut: apabila semakin kecil nilai rasio *SW* dan *SB* maka semakin baik, dengan demikian terdapat heterogenitas maksimum antar kelompok dan homogenitas maksimum dalam kelompok. Pengukuran kinerja dari data kategorik yang berbeda dengan data numerik yang diperoleh. Dengan melakukan pengukuran ANOVA yang didapatkan dari data kategorik dapat digunakan tabel kontingensi yang memiliki hasil ekuivalen.

Menurut Alvionita (2017), jika diperoleh *n* data yang diamati dengan *nk* adalah jumlah data yang diamati dari kategori ke-*k* dengan *k*=1,2,3,4,…,*K* dan $\sum\_{k=1}^{k}n\_{k}=n$. Jumlah data yang diamati dari kategori ke-*k* dan kelompok ke-*c* adalah *nkc*, dimana *c =* 1,2,3,…,*C* dengan *C* adalah jumlah pengelompokkan yang terbentuk, sehingga $n\_{c}=\sum\_{k=1}^{k}n\_{kc}$ adalah jumlah data yang diamati pada kategori ke-*k*. Untuk mengetahui jumlah data yang diamati dapat dihitung dengan rumus (13).

$$n=\sum\_{c=1}^{C}n\_{c}=\sum\_{k=1}^{K}n\_{k}=\sum\_{k=1}^{K}\sum\_{k=1}^{C}n\_{k}^{2} (13)$$

Menurut Dewi (2012) untuk merumuskan persamaan *SumeSquareeTotal* untuk variabel dengan data kategorik adalah sebagai berikut:

$$SST\_{kategorik}= \frac{n}{2}-\frac{1}{2n}\sum\_{k=1}^{K}n\_{kc }^{2} (14)$$

Rumus persamaan *SumeSquareeWithin-Group* adalah sebagai berikut:

$$SSW\_{kategorik}= \sum\_{c=1}^{c}\left(\frac{n\_{c}}{2}-\frac{1}{2n\_{c}}\sum\_{k=1}^{K}n\_{kc }^{2}\right) (15)$$

Rumus persamaan *SumaSquareaBetweenaGroup:*

$$SSB\_{kategorik}=\frac{1}{2} \left(\sum\_{c=1}^{c}\frac{1}{n}\sum\_{k=1}^{K}n\_{kc}^{2}\right)-\frac{1}{2n}\sum\_{k=1}^{K}n\_{kc }^{2} (16)$$

Selanjutnya rumus persamaan *MeaneSquareeTotal*, *MeaneSquareewithin* dan *meansesquareebetween* adalah sebagai berikut:

$$MST\_{kategorik}=\frac{SST}{(n-1)} (17)$$

$$MSW\_{kategorik}=\frac{SSW}{(n-C)} (18)$$

$$MSB\_{kategorik}=\frac{SSB}{(n-1)} (19)$$

Rumus persamaan simpanganebaku di kelompok (*SW*) dan simpanganebakueantar kelompoke(*SB*) data kategorik ditunjukkan sebagaieberikut:

$$S'\_{w}=\left[MSW\_{kategorik}\right]^{{1}/{2}} (20)$$

$$S'\_{B}=\left[MSB\_{kategorik}\right]^{{1}/{2}} (21)$$

Dari persamaan yang telah tersaji, maka kinerja metode pengelompokkanedata kategorik dan data numerik memiliki kesamaan yang menunjukkan semakin baik rasio *S’W* dan *S’B* semakin kecil, maka terdapat heterogenitas maksimum antar kelompok dan homogenitas maksimum dalam kelompok.

1. **METODEIPENELITIAN**

Data yang digunakan pada penelitian merupakan data sekunder yang didapatkan dari publikasi Kementrian Pendidikan dan Kebudayaan dan publikasi LembagaeLayanan Pendidikan Tinggi Wilayah VI (LLDIKTI VI). Pengolahan data menggunakan *software R studio* 4.1.1. Data yang diamati sebanyak 37 PTS aktif di Kota Semarang. Data menggunakan 4 variabel data numerik, yaitu: (1) Mutu SDM dan Mahasiswa (*Input*), (2) Pengelolaan kelembagaan (Proses), (3) Pencapaian Kinerja Jangka Pendek (*Output*), dan (4) Pencapaian Kinerja Jangka Panjang (*Outcome*). Sedangkan variabel data kategorik meliputi 3 variabel: (1) Bentuk PTS, (2) Akreditasi PTS, dan (3) Kinerja Penelitian PTS.

Berikut adalah tahapan pengelompokkan PTS dengan metode *ensembleeROCK*:

1. Semua variabel dibuat statistika deskriptif.
2. Variabel dibagi menjadi data numerik dan kategorik.
3. Metode hirarki dipakai untuk membuat klaster pada data numerik (*singleelinkage, completeelinkage, averageelinkage*) dengan jarak *Euclidean* sebagai ukuran jarak.
4. Menghitung *R-square* untuk menentukan total pengelompokkan optimum.
5. Nilai yang diberikan rasio *SW/SB* hasil pengelompokkan untuk menentukan kinerja pengelompokkan terbaik.
6. Metode ROCK dengan *threshold* (*θ*) yang telah ditentukan untuk melakukan klaster pada data kategorik
7. Hasil dari pengelompokkan metodeeROCK di masing-masing nilai *θ* digunakan untuk menentukan total pengelompokkan optimum dengan nilai rasio *S’W/S’B* minimum.
8. Hasil klaster pada tahap 3 dan 4 dilakukan kombinasi.
9. Penggunaan metode ROCK untuk pengklasteran hasil kombinasi
10. Hasil dari pengelompokkan dengan menggunakan metode ROCK dari masing-masing nilai *θ* digunakan untuk menentukan total pengelompokkan optimum yang didasarkan nilai rasio *S’W/S’B* minimum.
11. *Profiling* karakteristik setiap kelompok berdasar hasil pengelompokkan metode ROCK.
12. Menyimpulkan hasil pengolahan data dan penelitian.
13. **HASIL DAN PEMBAHASAN**
	1. **Karakteristik PTS di Kota Semarang**

**Tabel 1.** Karakteristik data Numerik

| **Variabel** | ***Mean*** |
| --- | --- |
| Mutu SDM dan Mahasiswa (*Input*) | 0,921 |
| Pengelolaan kelembagaan (Proses) | 1,908 |
| Pencapaian kinerja jangka pendek (*Output*) | 0,331 |
| Pencapaian kinerja jangka panjang (*Outcome*) | 0,326 |

Kualitas SDM dan Mahasiswa *(input)* PTS di Kota Semarang cukup tinggi, dengan rata-rata sebesar 0,921 dimana terdapat 22 PTS yang memiliki nilai diatas rata-rata. Kualitas pengelolaan kelembagaan (proses) PTS juga cenderung tinggi karena nilai dari 26 PTS berada di atas rata-rata. Kualitas capaian kinerja jangka pendek (*output*) dan capaian kinerja jangka panjang (*outcome*) cenderung rendah, hanya 13 PTS yang memiliki nilai kualitas capain kinerja jangka pendek diatas rata-rata dan 12 PTS yang memiliki nilai kualitas capaian kinerja jangka panjang diatas rata-rata. Kota Semarang didominasi oleh PTS berbentuk Sekolah Tinggi dengan presentase 54%, PTS berakreditasi B dengan presentase 68% dan kinerja penelitian PTS klaster Binaan dengan presentase 67%.

* 1. **Pengelompokkan PTS denganiMetodei*Ensemble*iROCK**

Hasil pengelompokkan PTS dengan metodei*ensemble*iROCK padaidatainumerik, PTS dikelompokkanimenjadin2esampaie5 kelompok menggunakan metodeehirarki, meliputi *singleelinkage, completeelinkage*, dan *averageelinkage*. Kelompok disebut optimum apabila menghasilkan kelompok dengan nilai yang naik sangat tajam dan memiliki kecenderungan stabil terhadap kelompok berikutnya.

**Gambar 1.** Nilai *R-Square* Bedasarkan Total Pengelompokkan pada MetodeeHirarkie*Agglomerative*



Untuk *single linkage* memiliki nilai optimum 3, karena ada perbedaan nilai yang signifikan antara klaster 2 dan 3, serta cenderung stabil pada klaster 3 dan seterusnya. Begitu juga untuk *completeelinkage* dan *averageelinkage* yang memiliki nilai optimum 3. Maka terpilih jumlah optimum klaster untuk masing-masing *linkage* adalah 3.

**Tabel 2.** Rasio Hasil Pengelompokkan Data Numerik

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metode** | **Jumlah Kelompok** | **Sw** | **SB** | **Rasio** |
| *SingleeLinkage* | 3 | 0.1855 | 0.9502 | 0.1952 |
| *CompleteeLinkage* | 3 | 0.1859 | 0.7454 | 0.2494 |
| *AverageeLinkage* | 3 | 0.1859 | 0.7454 | 0.2494 |

Dari tabel 2 dapat disimpulkan pengelompokkan data numerik dengan metode *singleelinkage* menghasilkan 3 kategori kelompok PTS dimana pengelompokkan yang baik dengan menggunakan metode hirarkie*agglomerative*. Kelompoke1 terdapat 31 PTS, Kelompok 2 terdapat 5 PTS, dan kelompok 3 dihasilkan ­1 PTS yang memiliki nilai terbaik.

Pada data kategorik pengelompokkan dilakukan dengan metode ROCK menggunakan nilai θ (*threshold*) antara 0 sampai 1. Nilai rasio S’W dan S’B untuk θ sebesar 0,1sampai 0,9 ditunjukkan oleh Tabel 3.

**Tabel 3.** Nilai Rasio Pengelompokkan Data Numerik

|  |  |
| --- | --- |
| ***Threshold (θ)*** | ***Ratio*** |
| 0,1 | 1.97 × 10-16 |
| 0,2 | 1.97 × 10-16 |
| 0,3 | 3.91 × 10-16 |
| 0,4 | 0.00 |
| 0,5 | 0.00 |
| 0,6 | 1.09 × 10-16 |
| 0,7 | 1.09 × 10-16 |
| 0,8 | 1.09 × 10-16 |
| 0,9 | 1.09 × 10-16 |

Pada tabel 3 menujukkan nilai rasio terendah menunjukkan pengelompokkan dengan nilai θ sebesar 0,4 dan 0,5. Dari hasil tersebut tidak dapat digunakan dengan alasan tidak sesuai dengan yang diharapkan, dikarenakan semua objek pengamatan ada pada kelompok yang sama. Maka diambil rasio terendah selanjutnya yaitu pengelompokkan dengan nilai rasio sebesar 1.09 × 10-16, dimana dimiliki oleh θ antara 0,6 sampai 0,9. Karena bernilai sama, maka dapat diambil salah satunya yaitu 0,6. Pada threshold (θ) 0,6 menghasilkan 4 kelompok optimum. Kelompok 1 terdiri atas 24 PTS, kelompok 2 terdiri atas 5 PTS, kelompok 3 terdiri atas 5 PTS, dan kelompok 4 terdiri atas 3 PTS.

Pengelompokkan data campuran yang menggunakan metodee*ensemble*eROCK dengan nilaieθ yang digunakan merupakan analisis pengelompokkan kategorik. Berikut Tabel 4 adalah nilai rasio *S’W* dan *S’B* untuk θ dengan nilai 0,1sampai 0,9.

**Tabel 4** Nilai Rasio Pengelompokkan Data Campuran

|  |  |
| --- | --- |
| **Threshold (θ)** | **Ratio** |
| 0,1 | 2.61 × 10-1 |
| 0,2 | 1.97 × 10-16 |
| 0,3 | 1.97 × 10-16 |
| 0,4 | 0.00 |
| 0,5 | 3.06 × 10-16 |
| 0,6 | 3.06 × 10-16 |
| 0,7 | 3.06 × 10-16 |
| 0,8 | 3.06 × 10-16 |
| 0,9 | 3.06 × 10-16 |

Dapat dilihat pada Tabel 4 diketahui nilai rasio terendah di pengelompokkane*ensemble* ROCK mendapatkan nilai 1.97×10-16. Karena nilai θ antara 0,2 dan 0,3 sama, maka dapat diambil salah satunya saja. Maka dapat disimpulkan bahwa dari hasil pengelompokkan data campuran dengan menggunakan metode *ensemble*eROCK dengan θ sebesar 0,2 diperoleh 3 kelompok kategori. Kelompok 1 terdapat 29 PTS, kelompok 2 terdapat 7 PTS, dan kelompok 3 terdapat 1 PTS.

* 1. ***Profiling/*karakteristik setiap kelompok**

**Tabel 5.** Perbandingan Nilai *Mean* Data Numerik

|  |  |  |  |
| --- | --- | --- | --- |
| **Variabel** |  | ***Mean*** |  |
| **Kelompok 1** | **Kelompok 2** | **Kelompok 3** |
| *Input**Proses**Output**Outcome* | 0,8321,7740,0010,078 | TerendahTerendahTerendahTerendah | 1,8442,6541,1571,033 | Tertinggi | 2,3513,1301,1131,550 | TertinggiTertinggiTertinggi |

Pada tabel 5 menghasilkan nilai *mean* data numerik di kolom kelompok 3 mengalami kecenderungan lebih baik dibanding karakteristik kelompok lain. Kelompok 1 terlihat memiliki nilai terendah di seluruh variabel yang terdiri dari 29 PTS yang memiliki kualitas kurang baik. Pada kelompok ini didominasi oleh PTS berbentuk Sekolah Timggi, berakreditasi B dan kinerja penelitian klaster Binaan. Kelompok 2 merupakan kelompok sedang yang terdiri dari 7 PTS yang didominasi oleh PTS berbentuk Universitas, berakreditasi B dan memiliki kinerja penelitian klaster Utama. Pada kelompok 3 terdiri atas PTS terbaik di Kota Semarang. PTS ini berbentuk Universitas berakreditasi A dengan kinerja penelitian klaster Utama.

1. **KESIMPULAN**

Kota Semarang didominasi oleh PTS berbentuk Sekolah Tinggi dengan presentase 54%, PTS berakreditasi B dengan presentase 68% dan kinerja penelitian PTS klaster Binaan dengan presentase 67%. Dari hasil pengelompokkan data campuran yang menggunakan metodeiensembleiROCK dengan θ bernilai 0,2 dihasilkan 3 kategori kelompok. Dengan rincian: (1) kelompok 1 terdapat 29 PTS; (2) kelompok 2 terdapat 7 PTS; dan (3) kelompok 3 terdapat 1 PTS. Kelompok 1 merupakan PTS dengan kualitas rendah, kelompok 2 merupakan PTS dengan kualitas sedang, dan kelompok 3 merupakan kelompok dengan kualitas tinggi. Universitas Katholik Soegijapranata (UNIKA) adalah satu-satunya PTS di kelompok 3, yang artinya UNIKA adalah PTS terbaik di Kota Semarang. Kelompok 2 terdiri dari UniversitaseIslameSultaneAgunge(Unissula), Universitase17eAgustuse1945 (Untag Semarang), UniversitaseSemarang (USM), Universitas Muhammadiyah Semarang (Unimus), Universitas Stikubank (Unisbank), UniversitaseDianeNuswantoroe(Udinus), dan UniversitasePGRI Semarange(UPGRIS). Lalu untuk PTS lain yang berada di kelompok 1 atau kelompok rendah harus meningkatkan kualitas PTS-nya terutama *output* dan *outcome* agar menjadi PTS yang lebih baik lagi.

**DAFTAR PUSTAKA**

Alvionita. (2017). *Metode Ensembel ROCK dan SWFM Untuk Pengelompokkan Data Campuran Numerik dan Kategorik Pada Kasus Aksesi Jeruk*. Surabaya: Institut Teknologi Sepuluh Nopember.

Ariska, N. (2017*). Analisis Cluster dengan Metode Ensemble ROCK untuk Data Berskala Campuran Katergorik dan Numerik (Kasus: Mahasiswa Aktif Program Studi Statistika FMIPA UNM).* Makassar: Universitas Negeri Makassar.

Buchta, C., & Hahsler, M. (2017). *Clustering for Business Analytics (Package cba)*

Bunkers, M. J., & James, R. M. (1996). *Definition of Climate Region in The Northern Plains Using an Objective Cluster Modification Technique*. Journal of Climate, 130-146.

Dewi, A. (2012). *Metode Cluster Ensemble Untuk Pengelompokkan Desa Perdesaan Di Provinsi Riau*. Surabaya: Institut Teknologi Sepuluh Nopember.

Dutta, M., Mahanta, A. K., & Pujari, A. K. (2005). *QROCK: A Quick of the ROCK Algorithm for Clustering of Categorical Data.* Proceedings of the 15 IEEE International Conference on Data Engineering.

Guha, S., Rastogi, R., & Shim, K. (2000). *ROCK: A Robust Clustering Algorithm for Categorical Attributes*. Proceedings of the 15th International Conference on Data Engineering

Hair, J. F., Black, W. C., Babin, J. B., & Anderson, E. R. (2001). *Multivariate Data Analysis (Seventh ed.).* New Jersey: Prentice Hall Inc.

He, Z., Xu, X. i., & Deng, S. (2005). *A Cluster Ensemble Method for Clustering Categorical Data. Information Fusion*, 143-151.

He, Z., Xu, X. i., & Deng, S. (2005). *Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach.* Departement of Computer Science and Engineering Harbin Institute of Technology.

Johnson, R. A., & Bhattacharyya, G. K. (2010). *Statistics Principle and Method (Sixth ed.)*. United State of America: John Wiley & Sons, Inc. Johnson, R. A., & Winchern, D. W. (2007). *Applied Multivariat Statistical Analysis (Sixth ed.)*. New Jersey: pearson Education.

[Kemdikbud] Direktorat Jenderal Pendidikan Tinggi. (2020). *Klasterisasi Pendidikan Tinggi*. Dipetik 12 Juni 2020, dari <https://klasterisasi-pt.kemdikbud.go.id/>

[Kemdikbud] Direktorat Jenderal Pendidikan Tinggi. (2020). *Klasterisasi Perguruan Tinggi Tahun 2020*. Dipetik 12 Juni 2020, dari http://lldikti6.id/wp-content/uploads/2020/09/KLASTERISASI-PT-2020.pdf

[Kemdikbud] LLDIKTI Wilayah VI. (2020). *Daftar Perguruan Tinggi*. Dipetik 12 Juni 2020, dari <https://dashboard-lldikti6.kemdikbud.go.id/>

Kementerian Negara Pendayagunaan Aparatur Negara. (2008). *Modul Pengukuran dan Analisis Kinerja di Lingkungan Instansi Pemerintah*. Jakarta

Kementerian Pendidikan dan Kebudayaan RI. Keputusan Menteri Pendidikan dan Kebudayaan Nomor 87 Tahun 2014 Tentang Akreditasi Program Studi dan Perguruan Tinggi.

Pemerintah Republik Indonesia. (2003). *Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional*. Jakarta.

Pemerintah Republik Indonesia. (2012). *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi*. Jakarta.

Pemerintah Republik Indonesia. (2014). *Undang-Undang Republik Indonesia Nomor 4 Tahun 2014 Tentang Penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi.* Jakarta

Sharma, S. (1996). *Applied Multivariate Technique*. New York: John Wiley and Sons, Inc.