

## METODE ENSEMBLE ROBUST CLUSTERING USING LINKS (ROCK) UNTUK PENGELOMPOKAN PERGURUAN TINGGI SWASTA (PTS) DI KOTA SEMARANG

Berliana Jannah<sup>1</sup>, Iut Tri Utami<sup>2</sup>, Arief Rachman Hakim<sup>3</sup>

Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*e-mail: [berlianajannah@gmail.com](mailto:berlianajannah@gmail.com)

DOI: 10.14710/j.gauss.12.3.445-452

### Article Info:

Received: 2023-02-02

Accepted: 2024-02-20

Available Online: 2024-02-26

### 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 ROCK ensemble method (Robust Clustering Using Links). 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 ( $\theta$ ) is 0.2. The results of each group are: 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) Perguruan Tinggi Negeri (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 ke depan maka calon mahasiswa mempertimbangkan beberapa aspek, diantaranya (1) mutu SDM dan mahasiswa (*input*), (2) pengelolaan kelembagaan (*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. Pengelompokan PTS di Kota Semarang 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 *cluster ensemble*, yaitu teknik pengelompokan yang menggabungkan hasil dari pengelompokan beberapa algoritma. Menurut He, et al (2005) dengan metode ini akan mendapatkan kelompok yang lebih baik.

## 2. TINJAUAN PUSTAKA

Statistika deskriptif adalah metode statistik yang berhubungan dengan pengumpulan data serta penyajian suatu informasi untuk memberikan 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 frekuensi relatif data kategorik.

Analisis *cluster* merupakan analisis multivariat yang bertujuan untuk pengelompokan  $n$  pengamatan ke dalam  $C$  kelompok 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) analisis *cluster* merupakan metode primitif yang tidak dibutuhkan adanya asumsi yang dapat digunakan dalam mengelompokkan data dengan alasan pengelompokan data berdasarkan kemiripan serta ketidakmiripan.

Pengertian dari metode dalam pengelompokan melalui data numerik dapat berdasarkan ukuran ketidakmiripan ataupun jarak. Sedangkan untuk ukuran ketidakmiripan yang sering digunakan ialah jarak *euclidean*, yaitu antar 2 pengamatan dan seberapa pengamatan sejumlah  $n$ . Berikut adalah rumus jarak *euclidean*.

$$d_{ij} = \sqrt{(x_i - x_j)'(x_i - x_j)}, i, j = 1, 2, \dots, n \text{ dan } i \neq j \quad (1)$$

$$\text{dengan } \mathbf{x}'_i = [x_{1i}, x_{2i}, x_{3i}, \dots, x_{mi}], \mathbf{x}'_j = [x_{1j}, x_{2j}, x_{3j}, \dots, x_{mj}].$$

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

### a) Single Linkage

*Single linkage* sebagai pengelompokan 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:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (2)$$

### b) Complete Linkage

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

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (3)$$

### c) Average Linkage

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

$$d_{(UV)W} = \frac{1}{N_{UV}N_W} \sum_q \sum_r d_{qr} \quad (4)$$

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

$$SST = \sum_{i=1}^{m_{numerik}} \sum_{j=1}^n (x_{ij} - \bar{x}_j)^2 \quad (5)$$

Nilai *Sum Square Within Group (SSW)* dihitung dahulu sebelum menghitung nilai *SSB* dengan rumus ditunjukkan pada persamaaan (6).

$$SSW = \sum_{h=1}^k \sum_{j=1}^{m_{numerik}} \sum_{i=1}^{n_h} (x_{ijh} - \bar{x}_{jh})^2 \quad (6)$$

Nilai SSB dirumuskan dengan persamaan (7).

$$SSB = SST - SSW \quad (7)$$

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

$$R_{square} = \frac{SSB}{SST} = \frac{[SST-SSW]}{SST} \quad (8)$$

*R-square* dapat diartikan sebagai ukuran pembeda antar kelompok dengan sekitaran dari angka 0 sampai 1. *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 antarkelompok.

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 pengelompokan 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 pengelompokan, sedangkan untuk pengamatan yang memiliki hubungan yang tinggi dijadikan 1 kelompok. Syarat-syarat pengelompokan algoritma ROCK akan berhenti ketika dalam keadaan (1) Total kelompok yang diinginkan terpenuhi, dan (2) Tidak adanya *link* antar kelompok.

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

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

$$sim(X_i, X_j) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|}, \quad i \neq j \quad (9)$$

$$i = 1, 2, 3, \dots, n \text{ dan } j = 1, 2, 3, \dots, n$$

$$X_i: \text{Himpunan pengamatan ke-}i, \text{ dengan } X_i = \{x_{1i}, x_{2i}, x_{3i}, \dots, x_{m_{kategorik}}\}$$

$$X_j: \text{Himpunan pengamatan ke-}j, \text{ dengan } X_j = \{x_{1j}, x_{2j}, x_{3j}, \dots, x_{m_{kategorik}}\}$$

#### 1. Menentukan *neighbors*

Dalam pengamatan  $X_i$  yang dinyatakan  $X_j$  sebagai *neighbors* apabila nilai  $sim(X_i, X_j) \geq \theta$ . *Threshold* ( $\theta$ ) merupakan parameter yang dapat ditentukan oleh peneliti sehingga bisa berfungsi untuk mengontrol kedekatan hubungan antar objek yang bernilai antara 0 sampai 1.

#### 2. Menentukan *Link*

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

#### 3. *Goodness Measure*

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

$$g(C_i, C_j) = \frac{link(C_i, C_j)}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}} \quad (10)$$

$link(C_i, C_j) = \sum_{X_i \in C_i, X_j \in C_j} link(X_i, X_j)$  merupakan jumlah *link* dari seluruh kemungkinan pasangan objek pada  $C_i$  dan  $C_j$ . Sedangkan  $n_i$  dan  $n_j$  menunjukkan total

anggota kelompok ke- $i$  dan kelompok ke- $j$  untuk  $f(\theta) = \frac{1-\theta}{1+\theta}$ , dengan  $\theta$  sebagai nilai *threshold* yang dipakai.

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

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

Kinerja hasil pengelompokan 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  $S_w$  dan  $S_B$ . Nilai simpangan baku atau *within* ( $SW$ ) menurut Bunker dan James (1996) ditunjukkan pada persamaan (11) berikut:

$$S_w = \frac{1}{C} \sum_{c=1}^C S_c \quad (11)$$

$$S_B = \left[ \frac{1}{C-1} \sum_{c=1}^C (\bar{x}_c - \bar{x})^2 \right]^{1/2} \quad (12)$$

Aturan yang digunakan dalam kinerja metode pengelompokan 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.

Menurut Alvionita (2017), jika diperoleh  $n$  data yang diamati dengan  $n_k$  adalah jumlah data yang diamati dari kategori ke- $k$  dengan  $k=1,2,3,4,\dots,K$  dan  $\sum_{k=1}^K n_k = n$ . Jumlah data yang diamati dari kategori ke- $k$  dan kelompok ke- $c$  adalah  $n_{kc}$ , dimana  $c = 1,2,3,\dots,C$  dengan  $C$  adalah jumlah pengelompokan 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).

$$n = \sum_{c=1}^C n_c = \sum_{k=1}^K n_k = \sum_{k=1}^K \sum_{c=1}^C n_{kc}^2 \quad (13)$$

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

$$SST_{kategorik} = \frac{n}{2} - \frac{1}{2n} \sum_{k=1}^K n_{kc}^2 \quad (14)$$

Rumus persamaan *Sum Square Within Group* adalah sebagai berikut:

$$SSW_{kategorik} = \sum_{c=1}^C \left( \frac{n_c}{2} - \frac{1}{2n_c} \sum_{k=1}^K n_{kc}^2 \right) \quad (15)$$

Rumus persamaan *Sum Square Between Group*:

$$SSB_{kategorik} = \frac{1}{2} \left( \sum_{c=1}^C \frac{1}{n} \sum_{k=1}^K n_{kc}^2 \right) - \frac{1}{2n} \sum_{k=1}^K n_{kc}^2 \quad (16)$$

Selanjutnya rumus persamaan *Mean Square Total*, *Mean Square within* dan *means square between* adalah sebagai berikut:

$$MST_{kategorik} = \frac{SST}{(n-1)} \quad (17)$$

$$MSW_{kategorik} = \frac{SSW}{(n-C)} \quad (18)$$

$$MSB_{kategorik} = \frac{SSB}{(n-1)} \quad (19)$$

Rumus persamaan simpangan baku di kelompok ( $SW$ ) dan simpangan baku antar kelompok ( $SB$ ) data kategorik ditunjukkan sebagai berikut:

$$S'_w = [MSW_{kategorik}]^{1/2} \quad (20)$$

$$S'_B = [MSB_{kategorik}]^{1/2} \quad (21)$$

Dari persamaan yang telah tersaji, maka kinerja metode pengelompokan data 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.

### 3. METODE PENELITIAN

Data yang digunakan pada penelitian merupakan data sekunder yang didapatkan dari publikasi Kementerian Pendidikan dan Kebudayaan dan publikasi Lembaga Layanan 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 pengelompokan PTS dengan metode *ensemble ROCK*:

1. Semua variabel dibuat statistika deskriptif.
2. Variabel dibagi menjadi data numerik dan kategorik.
3. Metode hirarki dipakai untuk membuat klaster pada data numerik (*single linkage, complete linkage, average linkage*) dengan jarak *Euclidean* sebagai ukuran jarak.
4. Menghitung *R-square* untuk menentukan total pengelompokan optimum.
5. Nilai yang diberikan rasio  $SW/SB$  hasil pengelompokan untuk menentukan kinerja pengelompokan terbaik.
6. Metode ROCK dengan *threshold* ( $\theta$ ) yang telah ditentukan untuk melakukan klaster pada data kategorik
7. Hasil dari pengelompokan metode ROCK di masing-masing nilai  $\theta$  digunakan untuk menentukan total pengelompokan 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 pengelompokan dengan menggunakan metode ROCK dari masing-masing nilai  $\theta$  digunakan untuk menentukan total pengelompokan optimum yang didasarkan nilai rasio  $S'W/S'B$  minimum.
11. *Profiling* karakteristik setiap kelompok berdasar hasil pengelompokan metode ROCK.
12. Menyimpulkan hasil pengolahan data dan penelitian.

### 4. HASIL DAN PEMBAHASAN

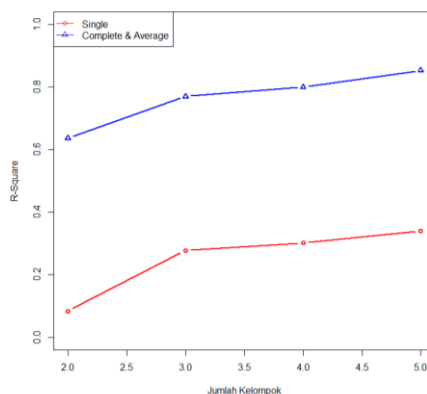
Tabel 1. Karakteristik data Numerik

Variabel	Mean
Mutu SDM dan Mahasiswa ( <i>Input</i> )	0,921
Pengelolaan kelembagaan ( <i>Proses</i> )	1,908
Pencapaian kinerja jangka pendek ( <i>Output</i> )	0,331
Pencapaian kinerja jangka panjang ( <i>Outcome</i> )	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 capaian 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%.

Hasil pengelompokan PTS dengan metode *ensemble* ROCK pada data numerik, PTS dikelompokkan menjadi 2 sampai 5 kelompok menggunakan metode hirarki, meliputi *single linkage*, *complete linkage*, dan *average linkage*. Kelompok disebut optimum apabila menghasilkan kelompok dengan nilai yang naik sangat tajam dan memiliki kecenderungan stabil terhadap kelompok berikutnya.



Gambar 1. Nilai *R-Square* Berdasarkan Total Pengelompokan pada Metode Hirarki *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 *complete linkage* dan *average linkage* yang memiliki nilai optimum 3. Maka terpilih jumlah optimum klaster untuk masing-masing *linkage* adalah 3.

Tabel 2. Rasio Hasil Pengelompokan Data Numerik

Metode	Jumlah Kelompok	$S_w$	$S_B$	Rasio
<i>Single Linkage</i>	3	0.1855	0.9502	0.1952
<i>Complete Linkage</i>	3	0.1859	0.7454	0.2494
<i>Average Linkage</i>	3	0.1859	0.7454	0.2494

Dari Tabel 2 dapat disimpulkan pengelompokan data numerik dengan metode *single linkage* menghasilkan 3 kategori kelompok PTS dimana pengelompokan yang baik dengan menggunakan metode hirarki *agglomerative*. Kelompok 1 terdapat 31 PTS, Kelompok 2 terdapat 5 PTS, dan kelompok 3 dihasilkan 1 PTS yang memiliki nilai terbaik.

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

Pada Tabel 3 menunjukkan nilai rasio terendah menunjukkan pengelompokan dengan nilai  $\theta$  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 pengelompokan dengan nilai rasio sebesar  $1.09 \times 10^{-16}$ , dimana dimiliki oleh  $\theta$  antara 0,6 sampai 0,9. Karena bernilai sama, maka dapat diambil salah satunya yaitu 0,6. Pada threshold ( $\theta$ ) 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.

Tabel 3. Nilai Rasio Pengelompokan Data Numerik

Threshold ( $\theta$ )	Ratio
0,1	$1.97 \times 10^{-16}$
0,2	$1.97 \times 10^{-16}$
0,3	$3.91 \times 10^{-16}$
0,4	0.00
0,5	0.00
0,6	$1.09 \times 10^{-16}$
0,7	$1.09 \times 10^{-16}$
0,8	$1.09 \times 10^{-16}$
0,9	$1.09 \times 10^{-16}$

Pengelompokan data campuran yang menggunakan metode *ensemble* ROCK dengan nilai  $\theta$  yang digunakan merupakan analisis pengelompokan kategorik. Berikut Tabel 4 adalah nilai rasio  $S'_W$  dan  $S'_B$  untuk  $\theta$  dengan nilai 0,1 sampai 0,9.

Tabel 4 Nilai Rasio Pengelompokan Data Campuran

Threshold ( $\theta$ )	Ratio
0,1	$2.61 \times 10^{-1}$
0,2	$1.97 \times 10^{-16}$
0,3	$1.97 \times 10^{-16}$
0,4	0.00
0,5	$3.06 \times 10^{-16}$
0,6	$3.06 \times 10^{-16}$
0,7	$3.06 \times 10^{-16}$
0,8	$3.06 \times 10^{-16}$
0,9	$3.06 \times 10^{-16}$

Dapat dilihat pada Tabel 4 diketahui nilai rasio terendah di pengelompokan *ensemble* ROCK mendapatkan nilai  $1.97 \times 10^{-16}$ . Karena nilai  $\theta$  antara 0,2 dan 0,3 sama, maka dapat diambil salah satunya saja. Maka dapat disimpulkan bahwa dari hasil pengelompokan data campuran dengan menggunakan metode *ensemble* ROCK dengan  $\theta$  sebesar 0,2 diperoleh 3 kelompok kategorik. Kelompok 1 terdapat 29 PTS, kelompok 2 terdapat 7 PTS, dan kelompok 3 terdapat 1 PTS.

Tabel 5. Perbandingan Nilai *Mean* Data Numerik

Variabel	Mean		
	Kelompok 1	Kelompok 2	Kelompok 3
<i>Input</i>	0,832	1,844	2,351
<i>Proses</i>	1,774	2,654	3,130
<i>Output</i>	0,001	1,157	1,113
<i>Outcome</i>	0,078	1,033	1,550

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 Tinggi, 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.

## 5. 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 pengelompokan data campuran yang menggunakan

metode ensemble ROCK dengan  $\theta$  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 Universitas Islam Sultan Agung (Unissula), Universitas 17 Agustus 1945 (Untag Semarang), Universitas Semarang (USM), Universitas Muhammadiyah Semarang (Unimus), Universitas Stikubank (Unisbank), Universitas Dian Nuswantoro (Udinus), dan Universitas PGRI Semarang (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 Ensemble ROCK dan SFWM untuk Pengelompokkan Data Campuran Numerik dan Kategorik pada Kasus Akses Jeruk*. Surabaya: Institut Teknologi Sepuluh Nopember.
- 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.
- 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. *Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach*. *Departement of Computer Science and Engineering Harbin Institute of Technology*.
- Pemerintah Republik Indonesia. 2012. *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi*. Jakarta.