

Pengelompokan Perguruan Tinggi Negeri di Indonesia Menggunakan Metode Hirarki dan K-Medoids dengan Ukuran Jarak Modifikasi Data Campuran

Nabila Puspa Hariani¹, Kariyam¹, Sekti Kartika Dini^{1*}

¹ Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta

*Corresponding e-mail : sektidini@uii.ac.id

ABSTRAK

Pendidikan merupakan salah satu sasaran pokok pemerintah dalam rangka meningkatkan kesejahteraan rakyat. Perguruan tinggi merupakan jenjang pendidikan setelah pendidikan menengah mencakup program pendidikan diploma, sarjana, magister, spesialis, dan doktor. Perguruan tinggi di Indonesia terdapat beberapa bentuk yaitu Politeknik, Sekolah Tinggi, Akademik, Universitas, dan Institut. Setiap perguruan tinggi mempunyai akreditasi universitas maupun akreditasi program studi. Akreditasi merupakan salah satu bentuk dari penilaian mutu dan kelayakan insitusi perguruan tinggi atau program studi yang dilakukan oleh organisasi atau badan mandiri di luar perguruan tinggi. Akreditasi tersebut dapat dikelompokkan dengan indikator lainnya untuk mengetahui pengelompokan Perguruan Tinggi Negeri di Indonesia pada tahun 2020. Pengelompokan tersebut berdasarkan indikator akreditasi dan 8 indikator lainnya diolah dengan menggunakan metode *Complete Linkage* dan *K-Medoids*. Manfaat dari penelitian ini ialah dapat dijadikan masukan dan motivasi bagi PTN di Indonesia untuk meningkatkan kualitasnya. Selain itu, hasil penelitian ini dapat berguna bagi calon mahasiswa baru dalam memilih Perguruan Tinggi untuk melanjutkan studi. Perhitungan dievaluasi menggunakan *Shilhouette Coefficient* dan *Davies Bouldin Index (DBI)* menghasilkan klaster optimal sebanyak 4 kelompok dengan metode *Complete Linkage* dan 2 kelompok dengan metode *K-Medoids*

Kata Kunci : Perguruan Tinggi Negeri, *Complete Linkage*, *K-Medoids*, *Shilhouette*, DBI

ABSTRACT

Education is one of the main targets of the government in order to improve the welfare of the people. Higher education is a level of education after secondary education which includes diploma, bachelor, master, specialist, and doctoral education programs. There are several forms of higher education in Indonesia, namely Polytechnics, Colleges, Academics, Universities, and Institutes. Each higher education has university accreditation and study program accreditation. Accreditation is a form of quality and feasibility assessment of higher education institutions or study programs carried out by independent organizations or bodies outside universities. This accreditation can be grouped with other indicators to find out the grouping of State Higher Education in Indonesia in 2020. The grouping is based on accreditation indicators and 8 other indicators are processed using the Complete Linkage and K-Medoids methods. The benefit of this research is that it can be used as input and motivation for PTNs in Indonesia to improve their quality. In addition, the results of this study can be useful for prospective new students in choosing higher education institutions to continue Calculations evaluated using the Shilhouette Coefficient and Davies Bouldin Index (DBI) resulted in optimal clusters of 4 groups using the Complete Linkage method and 2 groups using the K-Medoids method.

Keywords: State Higher Education, *Complete Linkage*, *K-Medoids*, *Shilhouette*, DB

I. PENDAHULUAN

Pendidikan merupakan proses berkelanjutan yang tidak pernah berakhir, sehingga dapat menghasilkan kualitas yang berkesinambungan, Pendidikan tinggi merupakan kelanjutan dari pendidikan menengah yang diselenggarakan dalam rangka menyiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan

akademik dan profesional dalam menerapkan dan mengembangkan ilmu pengetahuan, teknologi, dan kesenian [1]

Perguruan tinggi merupakan suatu penyelenggara pendidikan sebagai tingkat lanjut dari jenjang pendidikan menengah atas atau biasa disebut SMA. Hal ini sesuai dengan pengertian perguruan tinggi menurut UU No.20 tahun 2003 pasal 19 ayat 1 [2] yang menyatakan bahwa

“Perguruan tinggi merupakan jenjang pendidikan setelah pendidikan menengah mencakup program Pendidikan diploma, sarjana, magister, spesialis, dan doktor yang diselenggarakan oleh perguruan tinggi”

Perguruan tinggi di Indonesia terdapat beberapa bentuk yaitu Politeknik, Sekolah Tinggi, Akademi, Universitas, dan Institut. Pada setiap jenis bentuk tersebut biasanya memiliki sejumlah jurusan atau program studi yang ditawarkan. Setiap perguruan tinggi mempunyai akreditasi universitas maupun akreditasi program studi.

Akreditasi merupakan salah satu bentuk dari penilaian mutu dan kelayakan institusi perguruan tinggi atau program studi yang dilakukan oleh organisasi atau badan mandiri di luar perguruan tinggi. Bentuk penilaian mutu eksternal yang lain adalah berkaitan dengan akuntabilitas, pemberian izin, pemberian lisensi oleh badan tertentu [3]

Pada tahun 2020 Ditjen Dikti Kemendikbud mengumumkan hasil klusterisasi perguruan tinggi yang didasarkan atas indikator penilaian perguruan tinggi, yaitu indikator sumber daya manusia dan mahasiswa (bobot 20%), indikator pengelolaan kampus (bobot 25%), indikator pencapaian jangka pendek (bobot 25%), dan indikator pencapaian jangka panjang (bobot 30%). Hasil dari klusterisasi tersebut mengelompokkan perguruan tinggi di Indonesia menjadi 6 kelompok. Kelompok 1 yaitu perguruan tinggi peringkat 1 sampai 15 sehingga pada kelompok 1 menurut Dikti beranggotakan 15 perguruan tinggi. Kelompok 2 berisi 34 perguruan tinggi yaitu dari peringkat 16 sampai 50. Kelompok 3 berisi 97 perguruan tinggi yaitu dari peringkat 51 sampai 148. Kelompok 4 berisi 400 perguruan tinggi yaitu dari peringkat 149 sampai 549. Kelompok 5 berisi 1590 perguruan tinggi yaitu dari peringkat 550 sampai 2140. Kelompok 6 berisi 2141 perguruan tinggi yaitu dari peringkat 550 sampai 2140.

Analisis kelompok merupakan salah satu analisis multivariat yang bertujuan mempartisi objek ke dalam beberapa kelompok dengan karakteristik objek yang berada dalam satu kelompok relatif homogen daripada objek dalam kelompok yang berbeda. Ukuran kedekatan yang menjelaskan keserupaan antar objek merupakan prinsip dalam analisis kelompok. Ukuran kedekatan yang digunakan dalam pengelompokan objek disesuaikan dengan tipe data objek yang akan dikelompokkan [4]

Analisis kelompok terdiri dari dua jenis metode, yaitu hirarki dan non hirarki. Perbedaan mendasar dari kedua jenis metode tersebut ialah pada pembentukan jumlah kelompok. Contoh metode analisis kelompok hirarki ialah *Single Linkage Clustering*, *Complete Linkage Clustering*, dan *Average Agglomerative Clustering*. Sementara itu metode analisis kelompok non hirarki yang sering digunakan ialah *K-Means Clustering* dan *K-Medoids Clustering*.

Beberapa analisis kelompok yang sering digunakan adalah menggunakan metode hirarki dimana analisis pengelompokan data dilakukan dengan cara mengukur

jarak kedekatan pada setiap objek yang kemudian membentuk sebuah dendrogram atau diagram pohon. Selain itu, adapun metode non hirarki yaitu salah satunya adalah metode *K-Medoids* yakni dengan cara mengambil *centroid* menggunakan medoid sebagai titik acuan yang merupakan objek dalam sebuah klaster yang paling terpusat.

Adapun tujuan dari penelitian ini ialah melakukan pengelompokan Perguruan Tinggi Negeri (PTN) di Indonesia berdasarkan data yang tersedia di website PDDikti. Dengan demikian manfaat dari penelitian ini ialah dapat dijadikan masukan dan motivasi bagi PTN di Indonesia untuk meningkatkan kualitasnya. Selain itu, hasil penelitian ini dapat berguna bagi calon mahasiswa baru dalam memilih Perguruan Tinggi untuk melanjutkan studi. Data perguruan tinggi yang tersedia di website PDDikti mencakup variabel numerik dan variabel kategorik. Dengan demikian, pengelompokan data tersebut akan diolah menggunakan ukuran kedekatan jarak modifikasi data campuran serta akan dilakukannya evaluasi klaster dengan menggunakan koefisien *shilhouette* dan *Davies Bouldin Index (DBI)*. Perbedaan pengelompokan yang dilakukan pada penelitian ini dengan PDDikti ialah terletak pada variabel dan metode pengelompokan yang digunakan. Peneliti

II. STUDI LITERATUR

Penelitian mengenai pengelompokan data campuran sudah banyak dilakukan diantaranya tentang “*K-Medoid Clustering for Heterogeneous Datasets*” oleh [5]. Penelitian ini mengusulkan ukuran kemiripan baru dalam bentuk triplet untuk mencari jarak antara dua objek data dengan tipe campuran. Metode yang digunakan dalam penelitian ini ialah *K-Medoid Clustering* dengan memanfaatkan ukuran kemiripan dalam bentuk vektor. Hasil dari penelitian ini menunjukkan bahwa algoritma baru yang ditawarkan lebih unggul dari algoritma tradisional untuk data bertipe campuran. Hal ini berdasarkan validasi klaster menggunakan Indeks Purity dan Indeks Davies Bouldin.

Selanjutnya pada penelitian yang dilakukan oleh [6] mengenai “Penggunaan Analisis Two Step *Clustering* untuk Data Campuran”. Metode yang digunakan adalah metode *Two Step Clustering*. Penelitian ini menghasilkan kesimpulan analisis *two step clustering* mengelompokkan menjadi 3 kelompok model kendaraan. Variabel yang berpengaruh terhadap pembentukan klaster pertama adalah semua variabel kategorik dan kontinu. Klaster kedua yang berpengaruh adalah enam variabel yaitu satu variabel kategorik dan lima variabel kontinu. Sedangkan untuk klaster ketiga variabel yang berpengaruh ada delapan yaitu satu variabel kategorik dan tujuh variabel kontinu.

Penelitian selanjutnya dilakukan oleh [7] dengan judul “Metode *Cluster* Menggunakan Kombinasi Algoritma *Cluster K-Prototype* dan Algoritma Genetika Untuk Data Bertipe Campuran”. Metode yang digunakan adalah

metode *K-Prototype*. Penelitian ini menarik kesimpulan bahwa hasil perbandingan dari kedua metode tersebut ialah metode *K-Prototype* Genetika lebih baik dibandingkan metode *K-Prototype* tanpa Genetika.

Sementara itu, penelitian yang dilakukan oleh [8] “Implementasi Metode *K-Medoids Clustering* Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (*Hotspot*)”. Metode yang digunakan pada penelitian ini adalah metode *K-Medoids*. Kesimpulan yang dapat ditarik pada penelitian ini adalah kualitas *clustering* yang dihasilkan yaitu jumlah kelompok dan jumlah data mempengaruhi terhadap hasil *clustering*. Kelemahan dari penelitian ini adalah dianjurkan melakukan penambahan data dan parameter yang digunakan.

Penelitian lainnya dilakukan oleh [9] dengan judul “Analisis *K-Medoids Clustering* dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia”. Metode yang digunakan pada penelitian ini adalah metode *K-Medoids*. Kesimpulan dari penelitian ini adalah hasil kasus presentase data imunisasi Campak pada balita menggunakan *tools Rapidminer 5.3* diperoleh hasil yang sama dengan analisis perhitungan, dimana diperoleh 2 provinsi dengan *cluster* rendah yang menjadi pusat perhatian bagi pemerintah dalam melakukan sosialisasi dan pemerataan dalam pemberian imunisasi campak balita pada provinsi tersebut.

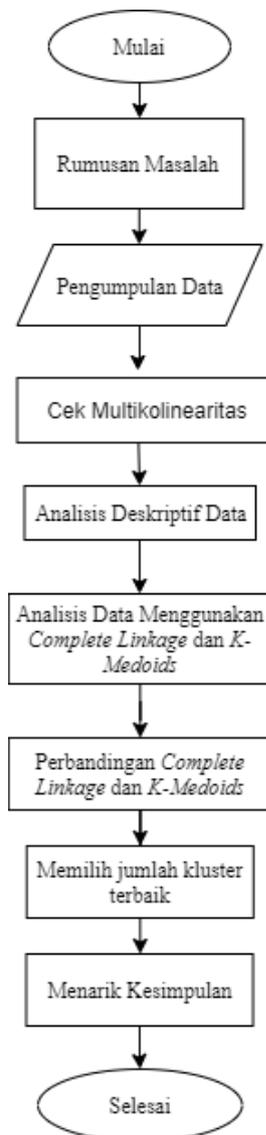
Penelitian lainnya oleh [10] yang berjudul “Implementation of *K-Medoids Clustering* for High Education Accreditation Data”. Metode yang digunakan pada penelitian ini adalah metode *K-Medoids*. Kesimpulan yang dapat ditarik pada penelitian ini adalah metode *K-Medoids* menghasilkan klaster terbaik.

III. METODOLOGI

Dalam penelitian ini, data yang digunakan diperoleh dari website PDDikti (<https://pddikti.kemdikbud.go.id>). Penelitian ini hanya menggunakan data Perguruan Tinggi Negeri (PTN) dengan jenis universitas dan institut. Hal ini karena universitas dan institut mempunyai kesamaan yang dekat yang membedakan hanyalah jika institut mempunyai satu spesifikasi keilmuan, berbeda dengan sekolah tinggi, akademi, dan politeknik yang dalam sistem pembelajarannya menerapkan praktik dalam bekerja. Dalam penelitian ini juga tidak mengikutsertakan data PTS karena keterbatasan data PTS

Sementara itu, analisis yang akan digunakan dalam mengelompokkan objek Perguruan Tinggi Negeri dengan menggunakan Metode Hirarki (*Complete Linkage*) dan *K-Medoids Clustering*.

Adapun beberapa tahapan penelitian yang dilakukan oleh peneliti terangkum pada Gambar 1



Gambar 1 : Diagram alir alur penelitian

Variabel yang digunakan dalam penelitian ini tersaji pada **Tabel 1**

A. Uji Multikolinearitas

Dalam analisis kelompok, multikolinearitas adalah suatu keadaan dimana ada hubungan linier sempurna atau hampir sempurna antar beberapa variabel. Menghitung nilai *VIF* (*Variance Inflation Factor*) merupakan salah satu cara dalam menguji multikolinearitas. Terjadinya multikolinearitas adalah jika memiliki nilai $VIF \geq 10$. Dimana nilai *VIF* dapat dihitung menggunakan **Persamaan 1**.

$$VIF = \frac{1}{1 - R_j^2} \quad (1)$$

Dengan R^2 merupakan koefisien determinasi yang nilainya dapat dihitung dengan **Persamaan 2**.

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 - (Y_i - \hat{Y})^2}{\sum_{i=1}^n (Y_i - \hat{Y})^2} \quad (2)$$

dengan,

Y_i : nilai data ke-i

\hat{Y} : nilai Y dari persamaan regresi

\bar{Y} : nilai rata-rata variabel Y

B. Ukuran Kedekatan antar Objek dengan Data Campuran

Berdasarkan ukuran antar objek menggunakan data campuran sebagai berikut:

- Ukuran kedekatan antar objek pada data numerik

$$d_{ij}numerik = \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{Max_k - Min_k} \quad (3)$$

dengan,

d_{ij} : jarak antara objek ke-i dan objek ke-j

x_{ik} : kumpulan variabel pada objek ke-i

x_{jk} : kumpulan variabel pada objek ke-j

k : banyaknya variabel ($x_1, x_2, x_3, \dots, x_p$)

Max_k : nilai maksimum variabel ke-k

Min_k : nilai minimum variabel ke-k

Berdasarkan ukuran jarak data numerik, maka digunakan ukuran jarak manhattanyang dimodifikasi

normalisasi *max-min* pada suatu variabel-k dengan nilai minimum pada suatu variabel-k

- Ukuran kedekatan antar objek pada data ordinal

$$d_{ij}ordinal = \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{Max_k - Min_k} \quad (4)$$

dengan,

d_{ij} : jarak antara objek ke-i dan objek ke-j

x_{ik} : kumpulan variabel pada objek ke-i

x_{jk} : kumpulan variabel pada objek ke-j

k : banyaknya variabel ($x_1, x_2, x_3, \dots, x_p$)

Max_k : nilai maksimum variabel ke-k

Min_k : nilai minimum variabel ke-k

Berdasarkan ukuran jarak data ordinal, maka digunakan ukuran jarak manhattanyang dimodifikasi normalisasi *max-min* pada suatu variabel-k dengan nilai minimum pada suatu variabel-k

Tabel 1 : Variabel Penelitian

No	Variabel	Notasi	Skala Data	Keterangan
1	Akreditasi	Akreditasi	Ordinal	1: A 2: Baik Sekali 3: Baik 4: B 5: Belum Terakreditasi
2	Jenis PTN	JenisPTN	Nominal	1: Universitas 2: Institut
3	Persentase Dosen Tidak Tetap	PDTT	Interval	$PDT = \frac{Jumlah\ Dosen\ Tidak\ Tetap}{Total\ Seluruh\ Dosen} \times 100\%$
4	Persentase Lektor Dosen Tetap	PL	Interval	$PL = \frac{Jumlah\ Dosen\ Tetap\ Lektor}{Total\ Dosen\ Tetap} \times 100\%$
5	Rasio Dosen Tetap terhadap Mahasiswa	RDTM	Interval	$RDTM = \frac{Jumlah\ Dosen\ Tetap}{Jumlah\ Mahasiswa}$
6	Persentase Dosen Kualifikasi S3	PS3	Interval	$PS3 = \frac{Jumlah\ Dosen\ Tetap\ Kualifikasi\ S3}{Total\ Dosen\ Tetap} \times 100\%$
7	Umur Perguruan Tinggi (Tahun)	Umur	Interval	
8	Persentase Akreditasi Program Studi Unggul (A)	PPA	Interval	$PPA = \frac{Jumlah\ Program\ Studi\ Unggul\ (A)}{Total\ Program\ Studi} \times 100\%$
9	Persentase Guru Besar (Profesor)	PP	Interval	$PP = \frac{Jumlah\ Dosen\ Tetap\ Profesor}{Total\ Dosen\ Tetap} \times 100\%$

- Perhitungan jarak data nominal
Perhitungan jarak data nominal menggunakan jarak *hamming*. Jarak *hamming* yaitu jarak biner antara dua nilai atribut x dan y menggunakan **Persamaan 5** [5]

$$d_{ij}nominal = 0 \text{ untuk } i = j; \quad (5)$$

$$d_{ij}nominal = 1 \text{ untuk } i \neq j$$

Berdasarkan ukuran jarak data nominal, digunakannya ukuran jarak *hamming* yaitu jika objek-i dengan objek-j mempunyai kemiripan kategori maka diberi nilai 0, namun sebaliknya jika tidak ada kemiripan kategori diberi nilai 1.

- Perhitungan total jarak data campuran

$$d_{ij}campuran = d_{ij}numerik + d_{ij}ordinal + d_{ij}nominal \quad (6)$$

C. Metode Pengelompokan Hirarki

Pengelompokan hirarki dimulai dengan dua atau lebih dari objek yang mempunyai kesamaan paling dekat. Analisis ini dilakukan hingga kelompok membentuk semacam pohon, dimana ada hirarki (tingkatan) yang jelas antarobjek dari yang paling mirip sampai dengan paling tidak mirip. *Dendogram* digunakan untuk membantu memperjelas proses hirarki tersebut [11]. Terdapat dua teknik pengelompokan dalam analisis hirarki yaitu teknik

pembagian (*dissive*) dan teknik penggabungan (*agglomerative*).

Teknik *agglomerative* menggabungkan dua kelompok yang terdekat dan seterusnya hingga diperoleh satu kelompok yang berunsurkan semua objek. Untuk menggabungkan dua kelompok pada tahap awal, dimana tiap kelompok hanya terdiri atas satu objek, diperlukan ukuran ketidakmiripan antar objek. Bila suatu kelompok merupakan penggabungan dari beberapa kelompok-kelompok dengan ukuran ketidakmiripan terkecil digabungkan menjadi kelompok yang baru.

Misalkan d_{uv} merupakan ukuran ketidakmiripan antara kelompok ke- u dengan kelompok ke- v dan $d_w(u,v)$ merupakan ukuran ketidakmiripan antara kelompok ke- w dengan kelompok (u,v) yang merupakan penggabungan antara kelompok ke- u dengan kelompok ke- v , maka beberapa teknik pengelompokan antara kelompok dinyatakan sebagai berikut [12]:

a. *Single Linkage*

Prosedur ini didasarkan pada jarak terkecil atau jarak terdekat. Jika dua objek terpisah oleh jarak yang pendek maka kedua objek tersebut digabung menjadi satu kelompok dan demikian seterusnya. Ukuran jarak yang digunakan adalah sebagai **Persamaan 7**

$$d_{w(u,v)} = \min(d_{wu}, d_{wv}) \quad (7)$$

b. *Complete Linkage*

Prosedur ini berlawanan dengan metode *single linkage* yang mengelompokkan berdasarkan jarak terbesar/terjauh. Ukuran jarak yang digunakan adalah sebagai **Persamaan 8**

$$d_{w(u,v)} = \max(d_{wu}, d_{wv}) \quad (8)$$

c. *Average Linkage*

Prosedur ini hampir sama dengan *single linkage* maupun *complete linkage*, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu kelompok dengan jarak seluruh individu dalam kelompok yang lain. Dengan n_u dan n_v merupakan jumlah pengamatan dalam kelompok ke- u dan ke- v , ukuran jarak yang digunakan adalah pada **Persamaan 9**

$$d_{w(u,v)} = \frac{n_u}{n_u+n_v} d_{wu} + \frac{n_v}{n_u+n_v} d_{wv} \quad (9)$$

D. Metode K-Medoids

Algoritma *K-Medoids* atau sering disebut juga dengan algoritma PAM (*Partitioning Around Medoid*) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw dimana merupakan algoritma yang mirip dengan *K-Means* karena kedua algoritma *partitional* yang memecah dataset menjadi kelompok-kelompok. Perbedaan dari algoritma *K-Means* dengan algoritma *K-Medoids* terletak pada penentuan pusat kluster, dimana algoritma *K-Medoids* menggunakan objek data perwakilan (*medoids*) sebagai pusat kluster [13]. Algoritma *K-Medoids* digunakan untuk mengatasi kelemahan dari algoritma *K-Means* yang sangat sensitif terhadap pencilan (*outlier*) karena objek-objek ini sangat

jauh letaknya/karakteristiknya dari mayoritas data lainnya, sehingga jika dimasukkan ke dalam suatu kelompok semacam ini bisa mendistrorsi nilai rata-rata (*mean*) kelompok tersebut. Adapun algoritma *K-Medoids* dijelaskan sebagai berikut:

- a. Pilihlah secara acak k objek pada sekumpulan n objek sebagai *medoid*
- b. Ulangi:
 1. tempatkan objek non-*medoid* ke dalam kelompok yang paling dekat dengan *medoid*
 2. secara acak pilih sebuah objek non-*medoid* sebagai kandidat *medoid* baru
 3. hitung total *cost* (jumlah jarak masing-masing objek ke *medoid* masing-masing)
 4. hitung total simpangan (S) dengan menghitung nilai total jarak baru – total jarak lama
- c. Lakukan perulangan hingga tidak terjadi perubahan

E. Silhouette Coefficient

Silhouette coefficient merupakan metode evaluasi untuk menguji ketepatan sebuah kluster yang telah terbentuk dari proses *clustering* [14]. *Silhouette coefficient* memberikan kualitas visual objek dalam tiap kluster dengan memberikan informasi sesuai dengan jumlah kluster pada data set. Adapun tahapan perhitungan *Silhouette Coefficient* adalah sebagai berikut [15]:

1. Hitung rata-rata jarak dari suatu data menggunakan Persamaan 10 maka didapatkan rata-rata dengan cara memisalkan I terhadap semua data lain yang berada dalam satu kluster.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (10)$$

dengan,

$a(i)$ = perbedaan rata-rata objek (i) ke semua objek lain pada A

$d(i, j)$ = jarak antara data i dan j

A = kluster (banyaknya data pada kluster ke- a)

2. Hitung rata-rata jarak data i tersebut dengan semua data di kluster lain, kemudian ambil nilai terkecilnya dengan Persamaan 11
- 3.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (11)$$

dengan,

$d(i, C)$ = perbedaan rata-rata objek (i) ke semua objek lain pada C

C = jumlah data kluster lain selain kluster A atau kluster C tidak sama dengan kluster A

4. Setelah menghitung $d(i, C)$ untuk semua C , maka diambil nilai terkecil menggunakan Persamaan 12

$$b(i) = \min_{c \neq A} d(i, C) \tag{12}$$

Klaster B yang mencapai minimum yaitu, $d(i, B) = b(i)$ disebut tetangga dari objek (i). ini adalah klaster terbaik kedua untuk objek (i).

5. Nilai *shilhouette coefficient* didefinisikan seperti pada persamaan 13

$$s(i) = \frac{(b(i)) - a(i)}{\max a(i), b(i)} \tag{13}$$

Tabel 2 menampilkan rentang nilai *Shilhouette Coefficient*

Tabel 2 : Nilai *Shilhouette Coefficient*

Nilai <i>Shilhouette Coefficient</i>	Struktur
$0.7 < SC \leq 1$	Struktur Kuat
$0.5 < SC \leq 0.7$	Struktur Sedang
$0.25 < SC \leq 0.5$	Struktur Lemah
$SC \leq 0.25$	Tidak Terstruktur

Nilai hasil perhitungan menggunakan *shilhouette coefficient* terletak pada nilai -1 sampai dengan 1. Nilai rata-rata *shilhouette coefficient* dalam suatu klaster dari setiap objek merupakan suatu ukuran untuk menunjukkan seberapa dekat kemiripan data dikelompokkan dalam satu klaster tersebut. Nilai rata-rata *shilhouette coefficient* semakin mendekati 1 maka semakin baik pengelompokannya dalam satu klaster, namun jika sebaliknya jika mendekati nilai -1 maka semakin buruk pengelompokan didalam satu klaster.

F. Davies Bouldin Index (DBI)

DBI merupakan salah satu evaluasi internal *cluster* untuk mengetahui baik atau tidak hasil dari pengelompokan yang ditinjau dari kohesi dan separasi [16]. Kohesi merupakan seberapa besar jarak antar data dengan tujuan mengukur seberapa besar kedekatan data yang berada pada klaster yang sama. Separasi merupakan nilai jarak untuk mengukur perbedaan data-data yang ada pada klaster yang berbeda. Langkah-langkah perhitungan *Davies-Bouldin Index* sebagai berikut:

1. *Sum of Square Within-Cluster (SSW)*
Sum Of Square Wpithin-Cluster dilakukan untuk mengetahui kohesi dalam sebuah klaster ke-i. Kohesi diartikan sebagai jumlah dari kedekatan atau kemiripan data terhadap titik pusat klaster dari sebuah klaster yang diikuti. Persamaan SSW yang digunakan adalah sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \tag{14}$$

dengan,

m_i = jumlah data dalam klaster ke-i

c_i = centroid klaster ke-i

$d(x_j, c_i)$ = jarak setiap data ke centroid i yang dihitung menggunakan jarak modifikasi

2. *Sum of Square Between-Cluster (SSB)*
Sum of Square Between-Cluster (SSB) dihitung dengan tujuan untuk mengetahui separasi atau jarak antar klaster. Berikut persamaan yang digunakan untuk menghitung SSB:

$$SSB_{i,j} = d(x_i, x_j) \tag{15}$$

Dimana $d(x_i, x_j)$ merupakan jarak antara data ke-i dengan data ke-j di klaster lain.

3. Rasio
Perhitungan rasio ($R_{i,j}$) bertujuan untuk mengetahui nilai perbandingan antara klaster ke-i dengan klaster ke-j dengan menghitung nilai rasio yang dimiliki oleh masing-masing klaster. Persamaan 16 menyajikan formula untuk menghitung rasio

$$R_{i,j,\dots,n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{ni,nj}} \tag{16}$$

dengan

SSW_i = *Sum Of Square Within Cluster* pada centroid i

$SSB_{i,j}$ = *Sum Of Square Between Cluster* data ke-i dengan j pada klaster yang berbeda

4. *Davies Bouldin Index (DBI)*
Nilai rasio yang diperoleh dengan menggunakan persamaan 16 digunakan untuk mencari nilai DBI menggunakan persamaan 17

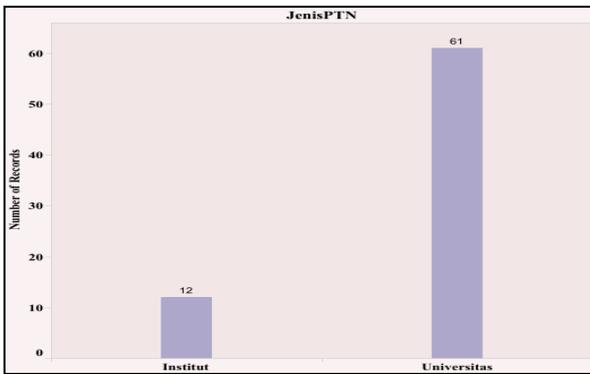
$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{i,j,\dots,n} \tag{17}$$

Dengan $R_{i,j,\dots,n}$ merupakan rasio dari nilai SSW dengan SSB pada persamaan 3.16. Dari perhitungan *Davies Bouldin Index (DBI)* dapat disimpulkan bahwa semakin kecil nilai *Davies Bouldin Index (DBI)* yang diperoleh (≥ 0) maka *cluster* tersebut semakin baik

IV. HASIL DAN PEMBAHASAN

A. Analisis Deskriptif

Statistika deskriptif dilakukan untuk mengetahui gambaran secara umum karakteristik untuk masing-masing variabel yang digunakan pada penelitian. Berikut merupakan gambaran umum masing-masing variabel.



Gambar 2 : Jumlah Perguruan Tinggi Negeri berdasarkan jenisnya

Berdasarkan Gambar 2 menginformasikan bahwa penelitian ini menggunakan 2 jenis perguruan tinggi. Ada sebanyak 73 Perguruan Tinggi Negeri (PTN) yang terdiri dari 61 universitas dan 12 institut.

Lokasi Perguruan Tinggi Negeri (PTN) yang digunakan pada penelitian ini tersebar di seluruh provinsi di Indonesia. Provinsi yang mempunyai perguruan tinggi paling banyak ada pada provinsi Jawa Timur yaitu sebesar 8 Perguruan Tinggi Negeri, dilihat dari ukuran tulisan yang sangat besar dan mempunyai ketebalan yang paling tebal. Selain itu, ada sebanyak 7 Perguruan Tinggi Negeri pada Provinsi Jawa Barat, sebanyak 6 Perguruan Tinggi Negeri pada Provinsi Jawa Tengah dan yang paling sedikit yaitu hanya ada 1 perguruan tinggi di tiap provinsi, seperti Provinsi Riau, Provinsi Kalimantan Barat, dan provinsi yang mempunyai ukuran besar tulisan yang kecil dan mempunyai ketebalan tulisan yang sangat tipis. Hal ini tersaji pada **Gambar 3**.



Gambar 3 : Lokasi Perguruan Tinggi

B. Uji Multikolinearitas

Asumsi yang perlu dipenuhi pada analisis kluster ialah tidak terjadinya multikolinearitas. Pada penelitian ini tidak terjadi multikolinearitas karena nilai VIF untuk semua variabel bernilai kurang dari 10.

C. Pengukuran Kedekatan Antar Objek

Kedekatan antar objek dilakukan dengan menggunakan jarak yang sudah termodifikasi. Berikut merupakan perhitungan jarak antar objek

1. Pengukuran Kedekatan Objek Data Nominal

Berikut hasil perhitungan jarak pada data nominal yang mempunyai ordo matriks 73 x 73.

$$d_{(i,j)nominal} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & \dots & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & \dots & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \end{pmatrix}$$

Berdasarkan matriks $d_{(i,j)nominal}$, perhitungan jarak antar objek pada jenis data nominal digunakan jarak *hamming*. Pada objek-1 (Universitas Airlangga) dengan objek-2 (Universitas Andalas), objek-3 (Universitas Brawijaya), objek-4 (Universitas Diponegoro) mempunyai kemiripan yang sama yaitu dengan jenis universitas, maka pada matriks tersebut diberi nilai 0 karena mempunyai kedekatan yang mirip. Sedangkan pada objek ke-8 (Institut Pertanian Bogor) mempunyai ketidakmiripan dengan objek-1 (Universitas Airlangga) maka pada baris ke-8 kolom-1 diberi nilai angka 1. Begitupun dengan nilai lainnya. Hasil dari perhitungan $d_{(i,j)nominal}$ merupakan total jarak keterdekatan antar objek pada data nominal yaitu variabel jenis Perguruan Tinggi Negeri (PTN) menggunakan jarak *hamming*.

2. Pengukuran Kedekatan Objek Data Ordinal

Berikut merupakan hasil perhitungan yang dimasukkan kedalam matriks data ordinal yang mempunyai ordo matriks 73 x 73.

$$d_{ordinal} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0.25 & 0.25 & 0.25 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0.25 & 0.25 & 0.25 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0.25 & 0.25 & 0.25 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0.25 & 0.25 & 0.25 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0.25 & 0.25 & \dots & 1 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0.25 & \dots & 1 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \dots & 0 \end{pmatrix}$$

Berdasarkan matriks $d_{A1,A73 ordinal}$ diberikan informasi bahwa pada objek-A1 (Universitas Airlangga) dengan objek-A2 (Universitas Andalas), objek-A3 (Universitas Brawijaya), objek-A4 (Universitas Diponegoro) mempunyai akreditasi yang sama yaitu A maka perhitungannya sesuai dengan d_{A1-A2} .

$$d_{A1,A2} = \frac{|1 - 1|}{5 - 1} = 0$$

3. Pengukuran Kedekatan Objek Data Numerik

Perhitungan untuk kedekatan objek data numerik dilakukan pada data PP, PL, RDTM, Umur, PPA, PDTT, dan PS3. Berikut ini merupakan matriks hasil total jarak data numerik:

$$d_{numerik} = \begin{pmatrix} 0 & 1.01 & 1.11 & 0.94 & 0.63 & 1.04 & 1.88 & 1.23 & \dots & 3.24 \\ 1.01 & 0 & 0.83 & 0.67 & 1.52 & 1.92 & 1.03 & 1.33 & \dots & 2.35 \\ 1.11 & 0.83 & 0 & 0.47 & 1.66 & 2.08 & 1.08 & 1.29 & \dots & 2.33 \\ 0.94 & 0.67 & 0.47 & 0 & 1.30 & 1.71 & 1.29 & 1.24 & \dots & 2.47 \\ 0.63 & 1.52 & 1.66 & 1.30 & 0 & 0.43 & 2.41 & 1.25 & \dots & 3.63 \\ 1.04 & 1.92 & 2.08 & 1.71 & 0.43 & 0 & 2.82 & 1.58 & \dots & 4.06 \\ 1.88 & 1.03 & 1.08 & 1.29 & 2.41 & 2.82 & 0 & 1.99 & \dots & 1.65 \\ 1.23 & 1.33 & 1.29 & 1.24 & 1.25 & 1.58 & 1.99 & 0 & \dots & 3.36 \\ \vdots & \ddots & \vdots \\ 3.24 & 2.35 & 2.33 & 2.47 & 3.63 & 4.06 & 1.65 & 3.36 & \dots & 0 \end{pmatrix}$$

Perhitungan jarak pada d_{A1-A73} numerik merupakan hasil dari total jarak seluruh variabel numerik yang disajikan dalam bentuk matriks

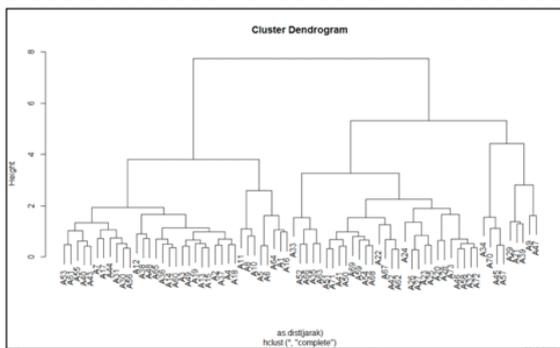
4. **Pengukuran Kedekatan Objek Data Campuran**

Setelah melakukan perhitungan pengukuran jarak pada data nominal, ordinal dan numerik, maka dilakukan penjumlahan dari seluruh matriks yang akan menghasilkan total kedekatan jarak antar objek data campuran sebagai berikut.

$$d_{campuran} = \begin{pmatrix} 0 & 1.01 & 1.11 & 0.94 & 0.63 & 1.04 & 1.88 & 2.23 & \dots & 4.24 \\ 1.01 & 0 & 0.83 & 0.67 & 1.52 & 1.92 & 1.03 & 2.33 & \dots & 3.35 \\ 1.11 & 0.83 & 0 & 0.47 & 1.66 & 2.08 & 1.08 & 2.29 & \dots & 3.33 \\ 0.94 & 0.67 & 0.47 & 0 & 1.30 & 1.71 & 1.29 & 2.24 & \dots & 3.47 \\ 0.63 & 1.52 & 1.66 & 1.30 & 0 & 0.43 & 2.41 & 2.25 & \dots & 4.63 \\ 1.04 & 1.92 & 2.08 & 1.71 & 0.43 & 0 & 2.82 & 2.58 & \dots & 5.06 \\ 1.88 & 1.03 & 1.08 & 1.29 & 2.41 & 2.82 & 0 & 2.99 & \dots & 2.65 \\ 2.23 & 2.33 & 2.29 & 2.24 & 2.25 & 2.58 & 2.99 & 0 & \dots & 4.36 \\ \vdots & \ddots & \vdots \\ 4.24 & 3.35 & 3.33 & 3.47 & 4.63 & 5.06 & 2.65 & 4.36 & \dots & 0 \end{pmatrix}$$

D. Pengelompokan menggunakan Metode Complete Linkage

Proses analisis menggunakan metode *complete linkage* merupakan proses *agglomerative*. Pengelompokan tersebut merupakan proses *clustering* yang dilakukan setelah jarak antar variabel diukur dengan rumus jarak data campuran (triplet) yang dimodifikasi. Oleh karena itu, perlu dilakukannya perhitungan total pengukuran jarak yang telah diproses serta disajikan dalam matriks $d_{campuran}$. Proses pengelompokan dapat diilustrasikan dalam bentuk dendrogram pada Gambar 4.



4 : Dendrogram metode Complete Linkage

Gambar 4 merupakan sebuah ilustrasi proses terbentuknya pengelompokan Perguruan Tinggi Negeri menggunakan metode *complete linkage*. Teknis pengelompokan dimulai dari dua objek yang memiliki jarak terjauh, dapat dijelaskan sebagai berikut:

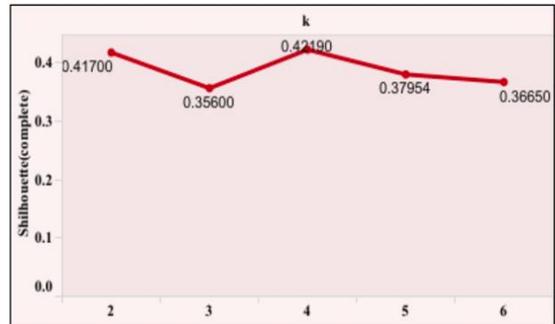
1. Tahap 1, objek A42 dengan A62 bergabung menjadi

satu kelompok menjadi kelompok 42. Objek A26 dengan A27 bergabung menjadi satu kelompok.

2. Proses penggabungan pengelompokan dilakukan hingga tahap 71 sampai semuanya bergerombol menjadi satu kelompok.

E. Evaluasi Metode Complete Linkage menggunakan Metode Shilhouette

Banyaknya kelompok yang terbentuk dikatakan baik jika nilai koefisien *shilhouette*-nya mendekati nilai 1. Gambar 5 menyajikan informasi terkait nilai koefisien *shilhouette* beserta banyaknya klaster yang terbentuk.



Gambar 5 : Grafik Shilhouette Metode Complete Linkage

Berdasarkan Gambar 5 diketahui bahwa dengan menggunakan metode *Complete Linkage* maka banyaknya kelompok yang baik digunakan adalah empat karena nilai koefisien *shilhouette*-nya paling mendekati satu dibandingkan nilai koefisien *shilhouette* lainnya.

F. Pengelompokan menggunakan Metode K-Medoids

Pada penelitian ini, akan dilakukan pengelompokan menjadi 2 kelompok. Langkah pertama yang dilakukan adalah dengan memilih *medoid* atau pusat klaster. Pusat klaster yang akan dipilih secara random adalah data ke-15 dan data ke-57. Data ke-15 merupakan Universitas Negeri Semarang sebagai pusat klaster 1 (*medoid 1*), sedangkan data ke-57 merupakan Institut Teknologi Kalimantan sebagai pusat klaster 2 (*medoid 2*). Langkah kedua adalah dengan menghitung jarak setiap objek terhadap pusat klaster menggunakan jarak triplet sebagai berikut.

Tabel 3 : Contoh jarak objek terhadap medoid

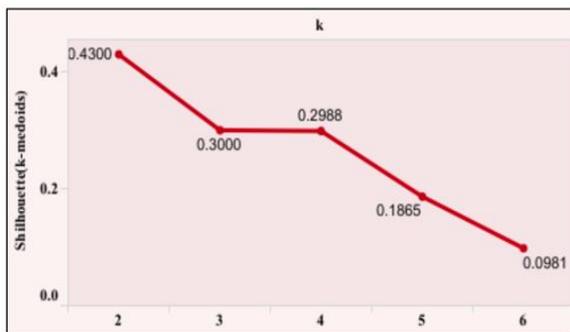
Anggota Kelompok 1	Medoid 1 (A15)	Anggota Kelompok 2	Medoid 1 (A57)
A1	1.85	A20	3.58
A2	0.74	A21	1.91
A3	0.51	A23	2.34
A4	0.55	A24	2.92
A5	1.74	A25	2.13

Berdasarkan Tabel 3 menginformasikan jarak objek terhadap pusat kelompok data. Pusat kelompok data terbagi menjadi dua. Jarak objek A1 dengan pusat kelompok A15 sebesar 1.85, begitupun dengan jarak objek A21 pada pusat kelompok A57 sebesar 3.58. Langkah ketiga adalah menentukan data non medoid secara random, namun pada

penelitian ini, objek selain pusat kelompok (*medoid*) dijadikan sebagai *non medoid*. Langkah keempat adalah menghitung nilai *cost*. Nilai *cost* yang diperoleh merupakan nilai paling minimum dari selisih antara nilai *cost* medoid dengan nilai *cost* non medoid yaitu sebesar 88.04. Pada penelitian ini, kelompok yang terbentuk sebanyak 2 kelompok dengan anggota kelompok 1 sebanyak 38 Perguruan Tinggi Negeri dan anggota kelompok 2 sebanyak 35 Perguruan Tinggi Negeri

G. Evaluasi Metode K-Medoids menggunakan Metode Shilhouette

Gambar 6 menyajikan informasi terkait nilai koefisien *shilhouette* beserta banyaknya kelompok yang terbentuk.



Gambar 6 : Grafik *Shilhouette* Metode *Complete Linkage* Berdasarkan Gambar 6 diketahui bahwa dengan menggunakan metode K-Medoids maka banyaknya kelompok yang baik digunakan adalah dua karena nilai koefisien *shilhouette*-nya paling mendekati satu dibandingkan nilai koefisien *shilhouette* lainnya.

H. Evaluasi Kluster menggunakan Davies Bouldin Index (DBI)

Nilai DBI untuk beberapa kelompok yang terbentuk tersaji pada Tabel 4

Tabel 4 : Nilai *Davies Bouldin Index*

Jumlah Kluster	DBI
2	0.79
3	1.16
4	4.47
5	3.40
6	3.41

Semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non negatif ≥ 0) maka kelompok yang terbentuk akan semakin baik. Berdasarkan Tabel 4, dapat

disimpulkan bahwa kelompok yang baik sebanyak 2

I. Perbandingan Evaluasi Kelompok

Berdasarkan jumlah kelompok optimal yang telah dihitung menggunakan evaluasi kluster *Shilhouette Coefficient* dan *Davies Bouldin Index* (DBI) maka diperoleh hasil yang terangkum pada **Tabel 5**.

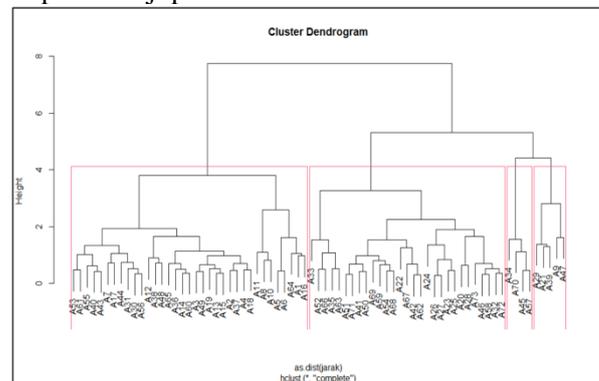
Tabel 5 : Perbandingan evaluasi kelompok

Jumlah Kelompok	<i>Shilhouette Coefficient Complete Linkage</i>	<i>Shilhouette Coefficient K-Medoids</i>	DBI
2	0.42	0.43	0.79
3	0.36	0.30	1.16
4	0.42	0.30	4.47
5	0.38	0.19	3.40
6	0.37	0.10	3.41

Berdasarkan Tabel 5 menginformasikan perbandingan nilai evaluasi kelompok. Pada Perguruan Tinggi Negeri di Indonesia pada Tahun 2020 akan dikelompokkan menjadi empat kelompok menggunakan metode *complete linkage* dan dua kelompok menggunakan metode *k-medoids*

J. Profilisasi Kelompok dengan Metode Complete Linkage

Analisis kelompok menggunakan *complete linkage* menghasilkan anggota kelompok yang dapat dilihat dengan *dendogram* atau biasa disebut diagram pohon. *Dendogram* yang terbentuk untuk pengelompokan Perguruan Tinggi Negeri di Indonesia sebanyak empat kelompok tersaji pada Gambar 7.



Gambar 7 : Dendogram *Complete Linkage* empat kelompok

Adapun profilisasi untuk setiap kelompok tersaji pada Tabel 6

Tabel 6 : Profilisasi kelompok dengan metode *Complete Linkage*

Kelompok	Anggota	Karakteristik Kelompok
1 (35 PTN)	Universitas Airlangga, Universitas Andalas, Universitas Brawijaya, Universitas Diponegoro,	Perguruan Tinggi Negeri di kelompok ini mempunyai rata-rata akreditasi lebih tinggi dibandingkan dengan

Kelompok	Anggota	Karakteristik Kelompok
	Universitas Gadjah Mada, Universitas Indonesia, Universitas Jendral Soedirman, Institut Pertanian Bogor, Institut Teknologi Bandung, Institut Teknologi Sepuluh Nopember, Universitas Negeri Malang, Universitas Negeri Semarang, Universitas Negeri Surabaya, Universitas Negeri Yogyakarta, Universitas Padjajaran, Universitas Pembangunan Nasional Veteran Jawa Timur, Universitas Pendidikan Indonesia, Universitas Sebelas Maret, Universitas Jember, Universitas Syah Kuala, Universitas Negeri Medan, Universitas Sumatera Utara, Universitas Negeri Padang, Universitas Riau, Universitas Sriwijaya, Universitas Lampung, Universitas Pendidikan Ganesha, Universitas Udayana, Universitas Tanjungpura, Universitas Lamhung Mangkurat, Universitas Mulawarman, Universitas Sam Ratulangi, Universitas Negeri Gorontalo, Universitas Hasanuddin, dan Universitas Negeri Makassar	klaster lain. Jika dilihat dari rata-rata tiap variabel, kelompok 1 mempunyai nilai variabel numerik rata-rata paling tinggi dari kelompok lain kecuali pada variabel rasio dosen tetap dan persentase lektor. Selain itu, pada anggota kelompok 1 seluruhnya mempunyai akreditasi institusi A yang paling tinggi dibandingkan dengan kelompok lain yang mempunyai variasi akreditasi pada setiap anggota klaster
2 (5 PTN)	Institut Seni Indonesia Yogyakarta, Institut Seni Indonesia Surakarta, Institut Seni Budaya Indonesia Bandung, Institut Seni Indonesia Padang Panjang, dan Institut Seni Indonesia Denpasar	Kelompok ini mempunyai kesamaan dengan kelompok 4, namun pada kelompok 2 berisi intitusi di bidang seni. Pada kelompok 2 mempunyai nilai rata-rata variabel RDTM dan persentase lektor paling tinggi dibandingkan dengan klaster lainnya.
3 (29 PTN)	Universitas Sultan Ageng Tirtayasa, Universitas Negeri Jakarta, Universitas Pembangunan Nasional Veteran Jakarta, Universitas Pembangunan Nasional Veteran Yogyakarta, Universitas Siliwangi, Universitas Singaperbangsa Karawang, Universitas Tidar, Universitas Trunojoyo, Universitas Malikussaleh, Universitas Samudera, Universitas Teuku Umar, Universitas Jambi, Universitas Bengkulu, Universitas Bangka Belitung, Universitas Mataram, Universitas Nusa Cendana, Universitas Timor, Universitas Palangka Raya, Universitas Borneo Tarakan, Universitas Negeri Manado, Universitas Tadulako, Universitas Sulawesi Barat, Universitas Sembilanbelas November Kolaka, Universitas Halu Oleo, Universitas Pattimura, Universitas Khairun, Universitas Cenderawasih, Universitas Mesamus Merauke, dan Universitas Papua	Pada kelompok ini, anggota seluruhnya adalah kumpulan universitas dimana nilai yang diperoleh cenderung lebih kecil dibandingkan kelompok 1 dan kelompok 2 jika dilihat dari nilai rata-rata variabel numerik. Pada kelompok ini, beberapa variabel numerik menduduki peringkat kedua kecuali variabel RDTM dan PPA.
4 (4 PTN)	Institut Seni Budaya Indonesia Aceh, Institut Teknologi Sumatera, Institut Teknologi Kalimantan, dan Institut Seni Budaya Indonesia Tanah Papua	Kelompok ini merupakan kelompok yang mempunyai nilai rata-rata variabel numerik paling rendah dibandingkan dengan kelompok yang lain

K. Profilisasi Kelompok dengan Metode K-Medoids

Analisis kelompok menggunakan Metode K-Medoids menghasilkan dua kelompok dengan profil setiap kelompok tersaji pada Tabel 7

Tabel 7 : Profilisasi kelompok dengan metode K-Medoids

Kelompok	Anggota	Karakteristik Kelompok
1 (38 PTN)	Universitas Airlangga, Universitas Andalas, Universitas Brawijaya, Universitas Diponegoro, Universitas Gadjah Mada, Universitas Indonesia, Universitas Jendral Soedirman, Institut Pertanian Bogor, Institut Seni Indonesia Yogyakarta, Institut Teknologi Bandung, Institut Teknologi Sepuluh Nopember, Universitas Negeri Malang, Universitas Negeri Semarang, Universitas Negeri Surabaya,	Perguruan Tinggi Negeri di kelompok ini mempunyai rata-rata akreditasi lebih tinggi dibandingkan dengan klaster kedua. Jika dilihat dari rata-rata tiap variabel, kelompok 1 mempunyai nilai yang lebih tinggi secara keseluruhan dibandingkan dengan kelompok 2 kecuali variabel RDTM

Kelompok	Anggota	Karakteristik Kelompok
	Universitas Negeri Yogyakarta, Universitas Padjajaran, Universitas Pembangunan Nasional Veteran Jawa Timur, Universitas Pendidikan Indonesia, Universitas Sebelas Maret, Universitas Negeri Jakarta, Universitas Jember, Universitas Syah Kuala, Universitas Negeri Medan, Universitas Sumatera Utara, Universitas Negeri Padang, Universitas Riau, Universitas Sriwijaya, Universitas Lampung, Institut Seni Indonesia Denpasar, Universitas Pendidikan Ganesha, Universitas Udayana, Universitas Tanjungpura, Universitas Lambung Mengkurat, Universitas Mulawarman, Universitas, Universitas Sam Ratulangi, Universitas Negeri Gorontalo, Universitas Hasanuddin, dan Universitas Negeri Makassar.	
2 (35 PTN)	Universitas Sultan Ageng Tirtayasa, Institut Seni Indonesia Surakarta, Universitas Pembangunan Nasional Veteran Jakarta, Universitas Pembangunan Nasional Veteran Yogyakarta, Universitas Siliwangi, Universitas Singaperbangsa Karawang, Universitas Tidar, Universitas Tronojoyo, Institut Seni Budaya Indonesia Bandung, Universitas Malikussaleh, Universitas Samudra, Institut Seni Budaya Aceh, Universitas Teuku Umar, Institut Seni Indonesia Padang Panjang, Universitas Jambi, Universitas Bengkulu, Institut Teknologi Sumatera, Universitas Bangka Belitung, Universitas Mataram, Universitas Nusa Cendana, Universitas Timor, Universitas Palangka Raya, Institut Teknologi Kalimantan, Universitas Borneo Tarakan, Universitas Negeri Manado, Universitas Tadulako, Universitas Sulawesi Barat, Universitas Sembilanbelas November Kolaka, Universitas Halu Oleo, Universitas Pattimura, Universitas Khairun, Universitas Seni Budaya Indonesia Tanah Papua, Universitas Cendrawasih, Universitas Musamus Merauke, dan Universitas Papua	Perguruan Tinggi Negeri di kelompok 2 mempunyai rata-rata akreditasi yang lebih rendah dibandingkan dengan kelompok 1. Jika dilihat dari rata-rata tiap variabel, kelompok 2 mempunyai nilai yang lebih rendah secara keseluruhan dibandingkan dengan kelompok 1 kecuali variabel RDTM yang lebih unggul dibandingkan dengan kelompok 1.

L. Perbandingan Hasil Klasterisasi Metode *Complete Linkage*, Metode *K-Medoids*, dan Dikti

Apabila dikaitkan dengan peringkat Perguruan Tinggi Negeri pada klasterisasi yang dilakukan oleh Dikti pada tahun 2020, maka tersaji pada Tabel 8.

Tabel 8 : Perbandingan hasil klasterisasi Metode *Complete Linkage*, Metode *K-Medoids* dan Dikti

No	PTN	Klaster Dikti	Klaster Metode <i>Complete Linkage</i>	Klaster Metode <i>K-Medoids</i>
1	IPB	1	1	1
2	UI	1	1	1
:	:	:	:	:
38	ISI Denpasar	4	2	1
39	UNRAM	2	3	2
40	UHO	3	3	2
:	:	:	:	:
73	ISBI Papua	5	4	2

Berdasarkan Tabel 8, diketahui bahwa klaster yang terbentuk oleh Dikti sebanyak enam klaster, namun pada

penelitian ini dengan menggunakan Metode *Complete Linkage* terbentuk empat kelompok. Kelompok 1 dan 3

pada penelitian ini sesuai dengan kelompok yang terbentuk pada kelompok Dikti, dimana kelompok 1 merupakan kelompok yang lebih tinggi dibandingkan dengan kelompok 3. Begitupun dengan kelompok 2 yang lebih tinggi dibandingkan dengan kelompok 4.

Sementara itu, dengan metode *K-Medoids* terbentuk dua kelompok. Kelompok 1 dan kelompok 2 sudah sesuai dengan urutan pada hasil klasterisasi oleh Dikti, dimana kelompok 1 merupakan kelompok dengan peringkat tinggi.

V. KESIMPULAN

Kesimpulan yang diperoleh pada penelitian ini ialah perbandingan hasil yang diperoleh antara metode *Complete Linkage* dan *K-Medoids* adalah terletak pada jumlah kelompok optimal. Berdasarkan metode *Complete Linkage* mempunyai jumlah kelompok optimal sebesar empat sedangkan jika menggunakan metode *k-medoids* jumlah kelompok optimalnya sebesar dua jika dilihat dari nilai koefisien *shilhouette* dan *Davies Bouldin Index*.

Kelompok yang terbentuk secara optimal dengan menggunakan metode *Complete Linkage*, profil dari kelompok optimal adalah pada kelompok 1 terdiri dari PTN yang mayoritas sudah berdiri sejak lama dengan akreditasi institusi A. Pada kelompok 1 ini merupakan kelompok paling tinggi diantara kelompok yang lainnya jika dilihat dari variabel indikator yang digunakan. Pada kelompok 2, karakteristik PTN mempunyai anggota yang seluruhnya berisi institut, kelompok ini mempunyai kesamaan dengan kelompok 4, namun pada kelompok 2 berisi institusi di bidang seni. Pada kelompok 2 mempunyai nilai rata-rata variabel RDTM dan persentase lektor paling tinggi dibandingkan dengan kelompok lainnya. Pada kelompok 4 merupakan kelompok yang mempunyai nilai paling rendah dibandingkan dengan kelompok yang lain.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Direktorat Penelitian dan Pengabdian Masyarakat (DPPM) UII atas Hibah Penelitian Unggulan berbasis Luaran yang telah diberikan.

REFERENSI

- [1] S. Yulawati, "Kajian implementasi tri dharma perguruan tinggi sebagai fenomena pendidikan tinggi di Indonesia," *Widya*, vol. 29, no. 318, pp. 28–33, 2012.
- [2] "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 20 TAHUN 2003 TENTANG SISTEM PENDIDIKAN NASIONAL," pp. 1–7, 2003.
- [3] BAN-PT, "Akreditasi Perguruan Tinggi Kriteria dan Prosedur 3.0," *Badan Akreditasi Nas. Perguru. Tinggi*, p. 18, 2019.
- [4] Kariyam, *Transformasi Data pada Analisis Kelompok*. Yogyakarta: Universitas Islam Indonesia, 2022.
- [5] S. Harikumar and P. V. Surya, "K-Medoid Clustering for Heterogeneous DataSets," *Procedia Comput. Sci.*, vol. 70, pp. 226–237, 2015, doi: 10.1016/j.procs.2015.10.077.
- [6] E. Charles and Mongi, "Penggunaan Analisis Two Step Clustering untuk Data Campuran Two Step Clustering Analysis for Combination Data," *JdC*, vol. 4, no. 1, 2015.
- [7] R. Nooraeni, S. Tinggi, and I. Statistik, "Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype Dan Algoritma Genetika Untuk Data Bertipe Campuran Cluster Method Using a Combination of Cluster K-Prototype Algorithm and Genetic Algorithm for Mixed Data," *J. Apl. Stat. Komputasi Stat.*, vol. 7, no. 2, pp. 17–17, 2015, [Online]. Available: <https://jurnal.stis.ac.id/index.php/jurnalasks/article/view/23>.
- [8] D. F. Pramesti, M. T. Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017.
- [9] S. Sundari, I. Sudahri Damanik, A. Perdana Windarto, H. Satria Tambunan, and A. Wanto, "K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita (Siti Sundari) Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, no. September, pp. 687–696, 2019.
- [10] Ghufuron, B. Surarso, and R. Gernowo, "Implementation of K-Medoids Clustering for High Education Accreditation Data," *J. Ilm. KURSOR*, vol. 10, no. 3, pp. 119–128, 2020.
- [11] J. Hair, R. Anderson, B. Babin, and W. Black, *Multivariate Data Analysis.pdf*, 7th ed. New Jersey: Prentice Hall Inc, 209AD.
- [12] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*, 6th ed. New Jersey: Pearson Prentice Hall, 2007.
- [13] N. K. Kaur, U. Kaur, and D. Singh, "K-Medoid Clustering Algorithm- A Review," vol. 1, no. 1, pp. 42–45, 2014.
- [14] M. Tanzil Furqon and L. Muflikhah, "Clustering the Potential Risk of Tsunami Using Density-Based Spatial Clustering of Application With Noise (DbSCAN)," *J. Enviromental Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 1–8, 2016, doi: 10.21776/ub.jeest.2016.003.01.1.
- [15] R. Handoyo, R. Rumani, and S. M. Nasution, "Perbandingan metode clustering menggunakan metode single linkage dan K-Means pada pengelompokan dokumen," *JSM STMIK Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014, [Online]. Available: <https://mikroskil.ac.id/ejurnal/index.php/jsm/article/view/161>.
- [16] A. Saikhu and Y. B. Gita, "Implementasi Deteksi Outlier pada Algoritma Hierarchical Clustering," *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 45–50, 2013.