

**PENERAPAN *HIERARCHICAL CLUSTERING MULTISCALE
BOOTSTRAP* PADA INDIKATOR PENDIDIKAN TINGKAT
SMA SEDERAJAT DI PROVINSI JAWA TIMUR**

SKRIPSI

**OLEH:
IMRO'ATUL KALAMIYAH
NIM. 19610110**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PENERAPAN *HIERARCHICAL CLUSTERING MULTISCALE
BOOTSTRAP* PADA INDIKATOR PENDIDIKAN TINGKAT
SMA SEDERAJAT DI PROVINSI JAWA TIMUR**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan
dalam Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Imro'atul Kalamiyah
NIM. 19610110**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

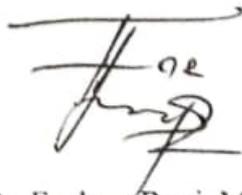
PENERAPAN *HIERARCHICAL CLUSTERING MULTISCALE BOOTSTRAP* PADA INDIKATOR PENDIDIKAN TINGKAT SMA SEDERAJAT DI PROVINSI JAWA TIMUR

SKRIPSI

Oleh
Imro'atul Kalamiyah
NIM. 19610110

Telah Disetujui untuk Diuji
Malang, 22 November 2023

Dosen Pembimbing I



Dr. Fachrur Rozi, M.Si.
NIP. 19800527 200801 1 012

Dosen Pembimbing II



Ach. Nashichuddin, M.A.
NIP. 19730705 200003 1 002

Mengetahui,
Ketua Program Studi Matematika



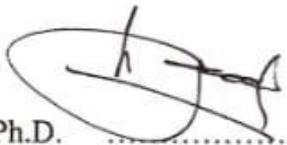
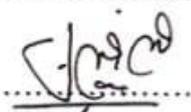
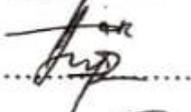
Dr. Ely Susanti, M.Sc.
NIP. 19741229 200012 2 005

PENERAPAN *HIERARCHICAL CLUSTERING MULTISCALE BOOTSTRAP* PADA INDIKATOR PENDIDIKAN TINGKAT SMA SEDERAJAT DI PROVINSI JAWA TIMUR

SKRIPSI

Oleh
Imro'atul Kalamiyah
NIM. 19610110

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)
Tanggal, 6 Desember 2023

Ketua Penguji	:	Prof. Dr. H. Turmudi, M.Si., Ph.D.	
Anggota Penguji 1	:	Ria Dhea Layla N. K., M.Si.	
Anggota Penguji 2	:	Dr. Fachrur Rozi, M.Si.	
Anggota Penguji 3	:	Ach. Nashichuddin, M.A.	

Mengetahui,
Ketua Program Studi Matematika


Dr. Ely Susanti, M.Sc.
NID. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertandatangan di bawah ini:

Nama : Imro'atul Kalamiyah

NIM : 19610110

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Penerapan *Hierarchical Clustering Multiscale*

Bootstrap pada Indikator Pendidikan Tingkat SMA

Sederajat di Provinsi Jawa Timur

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perilaku tersebut.

Malang, 6 Desember 2023
Yang membuat pernyataan



Imro'atul Kalamiyah
NIM. 19610110

MOTO

“Sesungguhnya bersama kesulitan itu ada kemudahan”

(Q.S Al-Insyirah: 6)

“Education is the most powerful weapon which you can use to change the world”

(Nelson Mandela)

PERSEMBAHAN

Alhamdulillah, dengan mengucapkan syukur kepada Allah *subhanahu wa ta'ala* skripsi ini penulis persembahkan kepada:

Allah *subhanahu wa ta'ala* Yang Maha Pengasih, ibu Ernalis Ekowati dan bapak Asikin selaku orang tua tersayang, kakak Bela Nabila, adik Muhammad Syifaul Qulub dan Aisha Shidqia Majidah.

Seluruh keluarga besar penulis terutama nenek dan kakek penulis dan kepada diri penulis sendiri.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakaatuh

Alhamdulillah, segala puji syukur penulis panjatkan ke hadirat Allah SWT yang telah memberikan limpahan rahmat, taufik dan hidayah sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Penerapan *Hierarchical Clustering Multiscale Bootstrap* pada Indikator Pendidikan Tingkat SMA Sederajat di Provinsi Jawa Timur”. Sholawat serta salam senantiasa terlimpahkan kepada Nabi junjungan yakni Nabi Muhammad SAW yang dengan tetap teguh dan gigih dalam memperjuangkan agama Islam sebagai agama pencerahan.

Penulis sadar bahwa penelitian ini tidak akan selesai tanpa adanya dukungan, bantuan, bimbingan dan nasihat dari berbagai pihak selama penyusunan ini. Pada kesempatan kali ini penulis menyampaikan rasa terima kasih kepada:

1. Prof. Dr. H. M. Zainuddin, MA., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, M.Sc., selaku ketua Program studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. H. Turmudi, M.Si., Ph.D., selaku ketua penguji dalam sidang skripsi yang telah memberikan ilmu pengetahuan dan saran yang bermanfaat bagi penulis.
5. Ria Dhea Layla Nur Karisma, M.Si., selaku anggota penguji 1 dalam sidang skripsi yang telah memberikan ilmu pengetahuan dan saran yang bermanfaat bagi penulis.
6. Dr. Fachrur Rozi, M.Si., selaku anggota penguji 2 sekaligus dosen pembimbing 1 yang telah memberikan berbagai nasihat, ilmu pengetahuan, arahan dan motivasi kepada penulis.
7. Achmad Nashichuddin, M.A., selaku anggota penguji 3 sekaligus dosen pembimbing 2 yang telah memberikan bimbingan, ilmu pengetahuan, arahan dan motivasi kepada penulis.
8. Seluruh dosen Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

9. Bapak Asikin dan Ibu Erna selaku orang tua penulis yang selalu mendoakan tanpa henti, memberikan dukungan material maupun moral sehingga penulis dapat menyelesaikan skripsi hingga akhir.
10. Kakak penulis, Bela Nabila dan adik-adik penulis, Asyif dan Aisha yang selalu mendoakan tanpa henti, memberi motivasi, dukungan dari awal hingga akhir.
11. Seluruh keluarga besar penulis yang tidak dapat disebutkan satu per satu, terutama nenek dan kakek penulis yang turut mendoakan, mendukung, dan meyakinkan penulis untuk segera menyelesaikan skripsi.
12. Seluruh teman-teman mahasiswa matematika angkatan 2019 yang telah memberikan bantuan dan dukungan dalam keadaan apapun.

Semoga Allah SWT. selalu memberikan balasan atas segala kebaikan dan keikhlasan yang telah diberikan kepada penulis dengan balasan yang terbaik. Pada akhirnya penulis berharap semoga adanya penelitian ini dapat bermanfaat bukan hanya bagi penulis melainkan juga kepada para pembaca untuk menambah ilmu pengetahuan dan juga wawasan.

Wassalamua'alaikum Warahmatullahi Wabarakatuh

Malang, 6 Desember 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	7
1.3 Tujuan Penelitian	7
1.4 Manfaat Penelitian	7
1.5 Batasan Masalahh	8
BAB II KAJIAN TEORI	9
2.1 Analisis <i>Cluster</i>	9
2.2 Deteksi <i>Outlier</i>	11
2.3 Uji Asumsi	12
2.4 Standardisasi Data.....	14
2.5 Metode Perhitungan Jarak.....	15
2.6 Metode <i>Clustering</i>	16
2.7 Validasi Analisis <i>Cluster</i> berdasarkan <i>Multiscale Bootstrap</i>	20
2.8 Indikator Pendidikan.....	25
2.9 Pendidikan dalam Alqur'an	28
2.10 Kajian Penelitian dengan Teori Pendukung	30
BAB III METODE PENELITIAN	32
3.1 Jenis Penelitian.....	32
3.2 Data dan Sumber Data	32
3.3 Instrumen Penelitian	32
3.4 Variabel Penelitian	33
3.5 Tahap Analisis Data	33
3.6 Diagram Alir Penelitian	36
BAB IV HASIL DAN PEMBAHASAN.....	37
4.1 Statistika Deskriptif Karakteristik Indikator Pendidikan SMA Sederajat Kabupaten/Kota di Provinsi Jawa Timur Tahun 2021	37
4.2 Deteksi <i>Outlier</i>	42
4.3 Uji Multikolinieritas.....	43

4.4	Standardisasi Data.....	45
4.5	Matriks Jarak Kedekatan <i>Euclidean</i>	45
4.6	Analisis <i>Agglomerative</i> Metode <i>Hierarchical Clustering</i> <i>Multiscale Bootstrap</i>	47
4.7	Penentuan Banyak <i>Cluster</i>	57
4.8	Karakterisasi Hasil <i>Cluster</i>	59
BAB V PENUTUP		64
5.1	Kesimpulan	64
5.2	Saran	65
DAFTAR PUSTAKA		66
LAMPIRAN		70
RIWAYAT HIDUP		80

DAFTAR TABEL

Tabel 3.1	Variabel Penelitian.....	33
Tabel 4.1	Statistik Deskriptif Data Indikator Pendidikan Tingkat SMA Sederajat di Provinsi Jawa Timur	37
Tabel 4.2	Uji Multikolinieritas Variabel APK, APS, APM, Pend. SMA Sederajat yang Ditamatkan, RLS, AMH.....	44
Tabel 4.3	Uji Multikolinieritas Variabel APK, APS, APM, Pend. SMA Sederajat yang Ditamatkan dan AMH	44
Tabel 4.4	Hasil Perhitungan AU, BP, v dan c dengan Metode <i>Multiscale Bootstrap</i>	56
Tabel 4.5	Nilai Koefisien dan Selisih Jarak <i>Complete Linkage</i>	58
Tabel 4.6	Jumlah <i>Cluster</i> dan Anggota dari Setiap <i>Cluster</i> dengan Metode <i>Hierarchical Clustering Multiscale Bootstrap</i>	59
Tabel 4.7	Hasil Perhitungan Nilai <i>Mean</i> pada Masing-Masing Variabel berdasarkan <i>Cluster</i>	60

DAFTAR GAMBAR

Gambar 2.1	Visualisasi Metode <i>Boxplot</i>	12
Gambar 2.2	Contoh <i>Dendrogram</i> Metode Hierarki	19
Gambar 2.3	Contoh <i>Dendrogram</i> Metode <i>Multiscale Bootstrap</i>	25
Gambar 3.1	Diagram Alir Tahap Analisis Data.....	36
Gambar 4.1	Diagram Variabel APK, APS, APM dan Pend. SMA Sederajat	39
Gambar 4.2	Diagram Variabel Rata-Rata Lama Sekolah.....	41
Gambar 4.3	Diagram Variabel Angka Melek Huruf	42
Gambar 4.4	Visualisasi <i>Boxplot</i> Data Indikator Pendidikan SMA Sederajat	43
Gambar 4.5	<i>Dendrogram Complete Linkage</i>	52
Gambar 4.6	Plot Diagnostik dari <i>Multiscale Bootstrap</i> untuk Langkah Penggabungan 1.....	55
Gambar 4.7	Hasil Akhir Visualisasi <i>Hierarchical Clustering</i> <i>Multiscale Bootstrap</i>	57
Gambar 4.8	Hasil <i>Dendrogram</i> Metode <i>Hierarchical Clustering</i> <i>Multiscale Bootstrap</i>	58

DAFTAR SIMBOL

R_j^2	: Koefisien determinasi hasil regresi dari variabel bebas X_j dengan variabel bebas lainnya
x_{ij}	: Nilai dari objek ke- i pada variabel ke- j
\bar{x}_j	: Nilai rata-rata dari variabel ke- j
s_j	: Nilai standar deviasi dari variabel ke- j
$d_{i,k}$: Jarak <i>euclidean</i> antara objek ke- i dengan objek ke- k
x_{kj}	: Nilai dari objek ke- k pada variabel ke- j
n	: Jumlah objek pengamatan
p	: Banyaknya variabel
$d_{(UV)W}$: Jarak antara <i>cluster</i> (UV) dan W
d_{UW}	: Jarak antara <i>cluster</i> U dan W
d_{VW}	: Jarak antara <i>cluster</i> V dan W
X_i	: Vektor data pengamatan ke- i
X_{i1}	: Nilai variabel X_1 dalam vektor pengamatan ke- i , $i = 1, 2, 3 \dots n$
X_{i2}	: Nilai variabel X_2 dalam vektor pengamatan ke- i , $i = 1, 2, 3 \dots n$
m	: Jumlah iterasi
$r = \sqrt{N'/N}$: Panjang urutan relatif dari replikasi <i>bootstrap</i>
N	: Jumlah data pengamatan
N'	: Jumlah data pengamatan <i>bootstrap</i>
v	: Jarak yang memiliki tanda (<i>signed distance</i>)
c	: Kelengkungan (<i>curvature</i>) dari batas
$\Phi(\cdot)$: Fungsi distribusi normal baku

DAFTAR LAMPIRAN

Lampiran 1. Data Indikator Pendidikan SMA Sederajat di Jawa Timur Tahun 2021	70
Lampiran 2. Hasil Standardisasi Data	71
Lampiran 3. Tabel Jarak <i>Euclidean</i>	72
Lampiran 4. Penggabungan <i>Cluster</i> Hierarki <i>Complete Linkage</i>	73
Lampiran 5. <i>Script</i> Proses Analisis <i>Hierarchical Clustering</i> <i>Multiscale Bootstrap</i>	78

ABSTRAK

Kalamiyah, Imro'atul. 2023. **Penerapan *Hierarchical Clustering Multiscale Bootstrap* pada Indikator Pendidikan Tingkat SMA Sederajat di Provinsi Jawa Timur.** Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: I) Dr. Fachrur Rozi, M.Si., II) Ach. Nashichuddin, M.A.

Kata Kunci: *Hierarchical Clustering Multiscale Bootstrap*, Pendidikan, Jawa Timur, Validitas *Cluster*.

Analisis *cluster* adalah salah satu dari banyak analisis dalam statistika yang dapat diterapkan dalam pengelompokan beberapa objek yang berkarakter serupa. Analisis *cluster* hanya memberikan satu solusi penyelesaian, yaitu didasarkan pada ukuran jarak kedekatan yang digunakan. Sehingga, penelitian ini menggunakan *hierarchical clustering multiscale bootstrap* untuk mengelompokan serta mengukur validitas setiap *cluster*, dengan memperhitungkan ketidakpastian dan stabilitas hasil *clustering*. Penelitian ini bertujuan untuk mengetahui hasil pengelompokan daerah-daerah yang memiliki kemiripan karakteristik indikator pendidikan dan mengetahui hasil validitas setiap *cluster*. Tahapan analisis yang dilakukan yaitu, melakukan perhitungan multikolinieritas dilanjutkan dengan menentukan jarak kedekatan menggunakan jarak *euclidean*, kemudian dilakukan pengelompokan menggunakan metode *hierarchical clustering multiscale bootstrap* untuk mendapatkan hasil pengelompokan dan validitas setiap *cluster*. Hasil penelitian diperoleh 2 *cluster* yang valid dengan nilai AU $p\text{-value} \geq 0,95$. *Cluster* 1 terdiri dari 26 Kabupaten/Kota dengan nilai AU = 0,98 dan *cluster* 2 terdiri dari 12 Kabupaten/Kota dengan nilai AU = 0,99.

ABSTRACT

Kalamiyah, Imro'atul. 2023. **Application of Hierarchical Clustering Multiscale Bootstrap on the Equivalent High School Education Indicators in East Java Province.** Thesis. Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisors: I) Dr. Fachrur Rozi, M.Si., II) Ach. Nashichuddin, M.A.

Keywords: Hierarchical Clustering Multiscale Bootstrap, Education, East Java, Cluster Validity.

Cluster analysis is one of many analyses in statistics that can be applied to grouping several objects of similar character. Cluster analysis provides only one solution to the solution, which is based on the size of the proximity distance used. Thus, this study uses hierarchical clustering multiscale bootstrap to group and measure the validity of each cluster, taking into account the uncertainty and stability of clustering results. This study aims to determine the results of grouping regions that have similar characteristics of educational indicators and determine the results of the validity of each cluster. The stages of the analysis carried out are to calculate multicollinearity followed by determining the proximity distance using Euclidean distance, then grouping using the hierarchical clustering multiscale bootstrap method to obtain the results of grouping and validity of each cluster. The results obtained two valid clusters with AU p-value ≥ 0.95 . Cluster 1 consists of 26 regencies/cities with AU value = 0.98 and cluster 2 consists of 12 regencies/cities with AU value = 0.99.

مستخلص البحث

الكلاميه، امرأة. ٢٠٢٣. تطبيق التجميع الهرمي متعدد المراحل بوتستراب (*Multiscale Bootstrap*) على مؤشرات التعليم الثانوي المكافئ في مقاطعة جاوى الشرقية. البحث الجامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج. المشرف: (١) الدكتور، فخر الرازي، الماجستير، (٢) أحمد ناصح الدين، الماجستير.

الكلمات الرئيسية: التجميع الهرمي متعدد المراحل بوتستراب (*Multiscale Bootstrap*)، التعليم، جاوى الشرق، صحة التجميع.

تحليل العنقودي هو من التحليل في الإحصاء الذي يمكن تطبيقه لتجميع عدة الكائنات التي لها طابع ممثلة. يستخدم التحليل العنقودي لحلّ مشهة واحدة باعتماد على المسافة القريبة. ويركز هذا البحث إلى استخدام تطبيق التجميع الهرمي متعدد المراحل بوتستراب (*Multiscale Bootstrap*) لتجميع وقياس صحة كل العنقودي، مع مراعاة عدم اليقين واستقرار نتائج التجميع. ويهدف هذا البحث إلى معرفة نتائج المناطق التي تشترك في خصائص مؤشرات التعليم ومعرفة نتائج صحة كل العنقودي. تتمثل مراحل التحليل في حساب التداخل الخطي المتعدد بتحديد مسافة القرب باستخدام المسافة الإقليدية، ثم يجري تجميع باستخدام تقنية التجميع الهرمي متعدد المراحل بوتستراب (*Multiscale Bootstrap*) للحصول على نتائج التجميع وصحة كل العنقودي. و يعرف نتائج هذا البحث للحصول على العنقودين صحيحين بالقيمة الاحتمالية $AU \leq 0,95$. و العنقودي الأول يتكون من ٢٦ منطقة أو مدينة بالقيمة $AU = 0,98$ ، والعنقودي الثاني يتكون من ١٢ منطقة أو مدينة بالقيمة $AU = 0,99$.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis *cluster* adalah salah satu dari banyak analisis dalam statistika yang dapat diterapkan dalam pengelompokan beberapa objek yang berkarakter serupa. Analisis *cluster* mengelompokkan objek yang paling mirip ke dalam kelompok yang sama sehingga sifat-sifat objek dalam satu *cluster* bersifat homogen, sedangkan sifat-sifat objek antar *cluster* lainnya bersifat heterogen (Hair, Black, Babin, & Anderson, 2010). Pada umumnya ada dua metode dalam analisis *cluster*, yaitu hierarki (*hierarchical clustering method*) dan non-hierarki (*non-hierarchical clustering method*). Metode hierarki memiliki dua jenis pendekatan, yaitu *agglomerative* dan *divisive*. Adapun *agglomerative* memiliki beberapa teknik, diantaranya *complete linkage*, *ward*, *single linkage* dan *average linkage*. Pada metode non-hierarki yaitu *K-means* (Supranto, 2004).

Metode hierarki merupakan metode dalam analisis *cluster* yang membentuk tingkatan seperti struktur pohon dikarenakan tahapan pengelompokannya dengan bertahap atau bertingkat. Pada analisis *cluster* dengan teknik pengukuran jarak hanya memberikan satu solusi dalam penyelesaiannya, yaitu didasarkan pada ukuran kedekatan jarak yang digunakan. Sehingga nilai signifikansi statistik dan ukuran akurasi dalam *clustering* tidak dapat diketahui dalam metode klasik ini (Anuraga, 2015). Oleh karena itu, dibutuhkan metode lain untuk mengatasi kelemahan pada analisis *cluster*; analisis *hierarchical clustering multiscale bootstrap* (HCMB) merupakan metode statistik yang digunakan untuk mengukur

ketidakpastian dalam *clustering* hierarki. Metode ini mengintegrasikan teknik *bootstrap* dengan algoritma *hierarchical clustering* untuk menghasilkan estimasi yang lebih akurat tentang struktur *clustering* dalam data dan stabilitas *cluster* (Suzuki & Shimodaira, 2004).

Analisis HCMB memberikan estimasi sejauh mana hasil *clustering* hierarki mungkin berbeda jika data diambil ulang (*resampling*), dengan melakukan repetisi *bootstrap* pada subset data pada skala yang berbeda, untuk memahami sejauh mana stabilitas hasil *clustering* hierarki terhadap variasi dalam data atau *resampling*. Dengan kata lain, ini membantu mengukur sejauh mana hasil *cluster* dapat dipercaya dan sejauh mana hasil *clustering* hierarki dapat dipertahankan jika data diambil ulang atau jika variasi dalam data berubah.

Pendidikan merupakan suatu hal penting yang berkaitan dengan peningkatan kualitas sumber daya manusia serta mendukung pembangunan suatu negara. Suatu negara dikatakan berkualitas jika di dalamnya terdapat sumber daya manusia yang berkualitas. Oleh karena itu, sumber daya manusia yang baik dan berkualitas akan menjadi pondasi pembangunan suatu negara (Dhira & Atika, 2019). Pendidikan dapat menjadi penentu kualitas sumber daya manusia. Semakin bagus pendidikan maka dapat menjadikan keteraturan dalam perbaikan kualitas sumber daya manusianya. Pendidikan adalah bentuk pembelajaran yang diberikan secara rapi dan sistematis serta jenjangnya terdiri dari pendidikan dasar, menengah dan tinggi. Tingkat pendidikan mempengaruhi perubahan mutu kehidupan sosial ekonomi penduduk. Adapun pendidikan yang layak diharapkan dapat mengubah pola pikir masyarakat agar lebih berkembang dan maju. Pemerintah telah lama menerapkan kebijakan guna meningkatkan kualitas sumber

daya manusia melalui akreditasi sekolah, standardisasi sekolah, dan wajib belajar 12 tahun (Badan Perencanaan Pembangunan Nasional, 2015).

Program wajib belajar adalah program minimal yang harus diselesaikan oleh warga negara dalam rangka meningkatkan potensi diri agar mampu hidup mandiri dalam lingkungan masyarakat maupun pendidikan tinggi. Pemerintah pusat maupun daerah sesuai wewenangnya wajib menyelenggarakan program wajib belajar. Pemerintah dapat menetapkan kebijakan guna meningkatkan jenjang pendidikan wajib belajar sampai pendidikan menengah atau dapat mengatur tergantung berdasarkan kondisi daerah masing-masing melalui peraturan daerah (Peraturan Pemerintah RI, 2008).

Kabupaten/Kota terbanyak dimiliki oleh Provinsi Jawa Timur dengan jumlah 38 Kabupaten/Kota. Jumlah yang besar mempersulit pemerintah untuk menyelesaikan masalah pendidikan, karena kondisi dan mutu pendidikan tidak merata. Penuntasan wajib belajar beralaskan pada indikator pendidikan salah satunya dengan melihat Angka Partisipasi Kasar (APK). Salah satu tujuan program wajib belajar pemerintah adalah meningkatkan angka partisipasi sekolah penduduk pada jenjang pendidikan menengah yang merata di seluruh Indonesia. Pemerintah telah membuat Rencana Strategis (Renstra) tahun 2020-2024 dengan target capaian indikator APK pada pendidikan menengah level Provinsi berturut-turut 86,18%; 88,39%; 88,39%; 92,80% dan 95% (Kementerian Kebudayaan dan Ristekdikti, 2022). Berdasarkan Survei Sosial Ekonomi Nasional (SUSENAS, 2021), Jawa Timur masih dalam angka 87,55% yang mana angka tersebut masih belum mencapai tujuan yang telah direncanakan oleh Kementerian Kebudayaan dan Ristekdikti. Beberapa Kabupaten/Kota seperti Kabupaten Lumajang, Kabupaten

Tulungagung, Kabupaten Probolinggo, Kabupaten Bangkalan dan Kabupaten Sampang masih memiliki nilai APK 50-60%. Selain APK, ada juga Angka Partisipasi Murni (APM) yang menjadi dasar indikator pendidikan untuk pemerataan kualitas pendidikan. Rata-rata APM di Jawa Timur hanya mencapai 62,63% di tahun 2021.

Pada tahun 2020 sekitar 24,21% penduduk Jawa Timur usia 10 tahun ke atas yang menempuh pendidikan hingga SMA sederajat, jumlah ini terbilang sangat minim dikarenakan sebanyak 27,02% penduduk telah menyelesaikan pendidikan SD/sederajat, artinya terbilang banyak penduduk usia sekolah yang tidak melanjutkan pendidikan hingga jenjang menengah (BPS, 2020). Berdasarkan informasi tersebut dapat dilihat bahwa terdapat disparitas indikator pendidikan di beberapa Kabupaten/Kota di Provinsi Jawa Timur.

Salah satu upaya yang dapat diterapkan demi mewujudkan pemerataan kondisi pendidikan wajib belajar 12 tahun di suatu wilayah adalah mengelompokkan wilayah-wilayah yang memiliki karakteristik yang sama, sehingga proses peningkatan dan pemerataan dapat sesuai dengan kondisi wilayah masing-masing. Oleh karena itu, pengelompokan masing-masing wilayah harus dilakukan, sehingga dapat dengan mudah membantu pemerintah mencapai keberhasilan dalam peningkatan dan pemerataan pendidikan. Teknik analisis yang dapat digunakan dalam mengelompokkan yaitu analisis *cluster*.

Terdapat ayat Alqur'an yang memberikan pemahaman tentang pendidikan, seperti dalam Alqur'an surah Al-'Alaq ayat 1-5:

اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ (١) خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ (٢) اقْرَأْ وَرَبُّكَ الْأَكْرَمُ (٣) الَّذِي عَلَّمَ بِالْقَلَمِ (٤) عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ (٥)

“(1) Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan, (2) Dia telah menciptakan manusia dari segumpal darah, (3) Bacalah, dan Tuhanmulah Yang Mahamulia, (4) Yang mengajar (manusia) dengan pena, (5) Dia mengajarkan manusia apa yang tidak diketahuinya” (Departemen Agama RI, 2006).

Surah Al-‘Alaq ayat 1-5 menekankan pentingnya membaca, mencari pengetahuan dan belajar sebagai bagian dari pendidikan. Membaca berarti berfikir secara teratur atau sistematis dalam mempelajari firman dan ciptaan-Nya. Manusia akan mampu menemukan konsep-konsep sains dan ilmu pengetahuan dengan berfikir. Tentu saja hal tersebut harus diawali dari membaca, karena membaca merupakan ilmu dasar untuk mendapatkan ilmu pengetahuan. Pemahaman dan pembelajaran yang diperoleh melalui proses ini dapat membantu dalam pengembangan pemikiran dan pemahaman yang baik. Manusia diingatkan Allah Swt. tentang keterbatasan pengetahuannya dan perlunya upaya aktif untuk belajar dan mencari pengetahuan (Sarwar, dalam Qutub, 2011).

Informasi mengenai pengelompokan digambarkan secara tersirat dalam Islam, seperti yang terkandung dalam Alqur’an surah Fathir ayat 32:

﴿ثُمَّ أَوْرَثْنَا الْكِتَابَ الَّذِينَ اصْطَفَيْنَا مِنْ عِبَادِنَا فَمِنْهُمْ ظَالِمٌ لِّنَفْسِهِ وَمِنْهُمْ مُّقْتَصِدٌ وَمِنْهُمْ سَابِقٌ بِالْخَيْرَاتِ
بِإِذْنِ اللَّهِ ۗ ذَٰلِكَ هُوَ الْفَضْلُ الْكَبِيرُ﴾

“Kemudian Kitab itu Kami wariskan kepada orang-orang yang Kami pilih di antara hamba-hamba Kami, lalu di antara mereka ada yang menzalimi diri sendiri, ada yang pertengahan dan ada (pula) yang lebih dahulu berbuat kebaikan dengan izin Allah. Yang demikian itu adalah karunia yang besar” (Departemen Agama RI, 2006).

Berdasarkan tafsir Jalalain, surah Fathir ayat 32 menjelaskan bahwa Alqur’an diberikan kepada tiga kelompok di antaranya: kelompok orang yang gegabah dalam mengamalkan. Kelompok kedua adalah orang pertengahan, yakni mereka yang berbuat kebajikannya setara dengan keburukannya. Kelompok ketiga adalah orang-orang yang cepat bertindak untuk kebaikan karena di samping

mengamalkan, juga mempelajari, mengajar dan membimbing orang lain untuk ikut mengamalkannya dengan izin Allah. Yang demikian itu diberikannya Alqur'an kepada mereka (itu adalah karunia yang sangat besar) (Al-Mahalli & Al-Suyuthi, 2007).

Penelitian terdahulu yang menggunakan analisis HCMB dengan metode hierarki *agglomerative* yaitu penelitian Safarina, Hayati, & Nasution (2019) dengan menerapkan HCMB berdasarkan Indeks Pembangunan Manusia di Kalimantan Timur tahun 2017. Hasil penelitian yang diperoleh yaitu terdapat empat *cluster* yang terbentuk namun, hanya tiga *cluster* yang valid. Metode *multiscale bootstrap* adalah metode yang dapat menunjukkan kevalidan setiap *cluster* yang terbentuk sebelumnya. Pada penelitian Ramly, Sudarmin & Poerwanto (2022) dengan mengaplikasikan tiga metode hierarki di antaranya *single linkage*, *average linkage* dan *complete linkage* kemudian melihat validitas dari hasil pengelompokan tersebut menggunakan *multiscale bootstrap*. Hasilnya, teknik yang terbaik adalah *single linkage* dengan empat *cluster* yang terbentuk valid. Metode HCMB merupakan metode pengelompokan hierarki yang bekerja dengan pendekatan *bootstrap resampling* untuk mendapatkan hasil akurasi *cluster*.

Berdasarkan penjelasan yang telah dipaparkan, penulis tertarik untuk melaksanakan sebuah penelitian dengan metode HCMB berdasarkan indikator pendidikan di Provinsi Jawa Timur tahun 2021. Penulis akan menyusun sebuah penelitian dengan judul "*Hierarchical Clustering Multiscale Bootstrap* pada Indikator Pendidikan Tingkat SMA sederajat di Provinsi Jawa Timur". Harapan penulis adalah agar penelitian ini dapat memberikan hasil pengelompokan terbaik dan gambaran kondisi pendidikan di setiap wilayah, sehingga dapat bermanfaat

untuk penyelesaian ketimpangan pendidikan di Kabupaten/Kota Provinsi Jawa Timur. Dengan demikian, pendidikan akan menjadi lebih baik dan dapat meningkatkan mutu sumber daya manusia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan, maka rumusan masalahnya adalah bagaimana hasil penerapan *hierarchical clustering multiscale bootstrap* dalam mengelompokkan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan serta mengetahui hasil validitas setiap *cluster*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian ini adalah untuk mengetahui hasil penerapan *hierarchical clustering multiscale bootstrap* dalam mengelompokkan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan serta mengetahui hasil validitas setiap *cluster*.

1.4 Manfaat Penelitian

Manfaat pada penelitian ini diharapkan dapat menyampaikan informasi kepada penulis sendiri, para pembaca dan instansi pemerintahan diantaranya:

1. Bagi penulis diharapkan dapat menambah pengetahuan mengenai penerapan analisis *cluster* dalam mengidentifikasi tingkat dan kondisi pendidikan SMA sederajat Kabupaten/Kota di Jawa Timur.
2. Bagi pembaca diharapkan dapat memberikan pengetahuan mengenai analisis *cluster*.

3. Bagi pemerintah diharapkan dapat memberikan informasi mengenai daerah-daerah yang membutuhkan perhatian khusus untuk meningkatkan kualitas pendidikan SMA sederajat.

1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah:

1. Ada enam dari tujuh indikator pendidikan, yaitu Angka Partisipasi Kasar (APK), Angka Partisipasi Murni (APM), Angka Partisipasi Sekolah (APS), Pendidikan SMA yang ditamatkan, Rata-rata Lama Sekolah (RLS), Angka Melek Huruf (AMH). Satu indikator pendidikan, yaitu Angka Buta Huruf (ABH) tidak digunakan karena tidak tersedianya data.
2. Menggunakan data indikator pendidikan Kabupaten/Kota di Jawa Timur tahun 2021.
3. Menggunakan pengukuran jarak *euclidean*.
4. Menggunakan metode hierarki *complete linkage*.

BAB II

KAJIAN TEORI

2.1 Analisis Cluster

Analisis *cluster* adalah suatu metode statistik yang digunakan untuk klasifikasi objek ke dalam suatu *cluster* berdasarkan karakteristik atau sifatnya. Penyelesaian *cluster* secara menyeluruh tergantung pada beberapa variabel yang digunakan sebagai dasar penilaian kesamaan. *Cluster* sendiri dapat didefinisikan sebagai sekelompok objek yang memiliki kesamaan dan dikelompokkan menjadi satu kelompok (Tan, Steinbach, & Kumar, 2006). Menurut Jain dan Dubes (1988) Algoritma *clustering* sering digunakan tidak hanya untuk klasifikasi data tetapi juga untuk pemodelan dan kompresi data. Selain itu, dapat mengumpulkan kelompok data dan digunakan untuk membangun model berdasarkan *cluster* yang telah dikelompokkan.

Pada dasarnya ciri suatu *cluster* yang baik yaitu (Wijaya & Budiman, 2016):

1. Homogenitas (kesamaan) maksimum antar anggota dalam satu kelompok
2. Heterogenitas (perbedaan) maksimum antar kelompok satu dengan yang lain.

Analisis *cluster* adalah teknik analisis multivariat yang bertujuan untuk mengelompokkan beberapa objek berdasarkan karakteristiknya. Analisis multivariat adalah pengembangan kata dari analisis univariat maupun bivariat. Kata multivariat berasal dari kata multi dan variat yang memiliki makna banyak variabel yang diamati. Analisis multivariat adalah analisis yang kompleks daripada analisis univariat. Salah satu analisis pada statistika yaitu analisis multivariat merupakan

metode pengolahan variabel lebih dari satu secara serentak. Analisis *cluster* mengelompokkan objek-objek yang paling dekat hubungannya dengan objek lain ke dalam *cluster* yang sama. Penyelesaian dalam *cluster* secara menyeluruh tergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan.

Langkah-langkah dalam analisis *cluster* adalah sebagai berikut (Supranto, 2004):

1. Merumuskan Masalah

Pertama yang harus dilakukan adalah dengan merumuskan masalah. Pemilihan variabel-variabel yang menggambarkan objek penting dilakukan ketika merumuskan masalah dalam analisis *cluster*.

2. Memilih Ukuran Jarak atau Similiaritas

Tujuan utama pengelompokan ialah mengklasifikasikan objek yang mirip dalam satu *cluster* atau kelompok, oleh karena itu pengukuran jarak diperlukan untuk mengetahui kemiripan dan perbedaan antar objek-objek yang diteliti. Objek yang memiliki jarak paling dekat akan memiliki kemiripan daripada objek yang memiliki jarak paling jauh

3. Standardisasi Data

Standardisasi data dilakukan jika masing-masing variabel yang digunakan berbeda satuan.

4. Memilih Prosedur *Clustering*

Pada tahapan ini objek yang memiliki kemiripan dan perbedaan dikelompokkan kemudian memilih metode *cluster* dan melakukan pengelompokan objek

5. Menginterpretasi *Cluster*

Tahap interpretasi dan memprofil *cluster* adalah pengkajian mengenai *centroids* atau nilai rata-rata setiap objek yang terdapat dalam *cluster* setiap objek. Nilai rata-rata digunakan untuk menguraikan setiap *cluster* dengan memberi suatu nama atau label pada setiap *cluster*

6. Uji Validitas dan *profiling*

Pada tahapan ini uji validitas digunakan untuk memastikan bahwa pengelompokan pada penyelesaian sebelumnya sudah menggambarkan populasi dan dapat digeneralisasikan ke objek yang lain. *Profiling* dilakukan untuk memberikan gambaran hasil *cluster* berdasarkan profil tertentu.

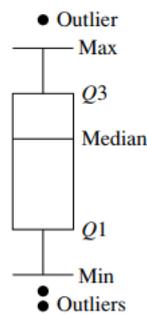
2.2 Deteksi *Outlier*

Sebuah data atau sekumpulan data pada umumnya mempunyai nilai-nilai pada setiap objek yang tidak terlalu berbeda dengan objek lain. Deteksi *outlier* adalah data *pre-processing* yang perlu dilakukan sebelum melakukan analisis *cluster*. Hal ini dilakukan karena analisis *cluster* bukan hanya sensitif terhadap variabel yang tidak relevan tetapi juga sensitif pada *outlier*.

Outlier dapat berupa pengamatan yang benar-benar menyimpang yang tidak mewakili populasi umum dan tidak representatif kelompok yang sebenarnya dalam populasi serta tidak signifikan dalam populasi. Adapun beberapa pembagian kasus pada *outlier*. Pertama, *outlier* mendistorsi struktur yang sebenarnya dan membuat *cluster* tidak mewakili struktur populasi yang aktual. Kedua, *outlier* dihilangkan sehingga hasil *cluster* akan lebih akurat mewakili segmen yang relevan dalam

populasi. Ketiga, *outlier* tetap harus disertakan dalam analisis *cluster*, meskipun kurang mewakili dalam sampel, karena *outlier* mewakili kelompok yang valid dan relevan. Maka dari itu, mendeteksi *outlier* perlu dilakukan pada tahap awal (Hair dkk., 2010).

Salah satu metode langsung untuk mendeteksi *outlier* dalam statistik juga dapat dalam bentuk visualisasi. Sebagai contoh, metode *boxplot* yaitu plot untuk menggambarkan secara grafik dari data numerik melalui lima ukuran di antaranya nilai observasi terkecil (*min*), kuartil bawah (Q1), median (Q2), kuartil atas (Q3) dan nilai observasi terbesar (*max*) (Han, Kamber, & Pei, 2011). Pada penelitian ini metode yang digunakan dalam pendeteksian *outlier* yaitu metode *boxplot*. Gambar 2.1 adalah contoh visualisasi *boxplot*.



Gambar 2.1 Visualisasi Metode *Boxplot*
(Sumber: Han, Kamber dan Pei (2011:555))

2.3 Uji Asumsi

Analisis *cluster* bukanlah teknik statistika inferensi yang mana parameter dari sampel dapat dikatakan mewakili populasi. Pada teknik selain analisis *cluster* persyaratan yang harus dilakukan yaitu uji normalitas, linieritas dan homoskedastisitas. Namun pada analisis *cluster* hanya perlu melakukan uji kecukupan sampel (*representativeness of the sample*) dan uji multikolinieiritas

(Hair dkk., 2010). Pada penelitian ini uji asumsi yang digunakan hanya uji multikolinieritas saja dikarenakan data yang akan digunakan pada penelitian ini adalah data populasi, sehingga uji kecukupan sampel dalam penelitian ini tidak dilakukan.

Uji multikolinieritas merupakan uji asumsi yang digunakan untuk memperlihatkan adanya korelasi antar dua atau lebih variabel. Untuk mengetahui ada tidaknya multikolinieritas terdapat beberapa cara yang dapat digunakan menurut Naes, Isaksson, Fearn, & Davies (2002) salah satu metodenya adalah dengan melihat variabel-variabel yang berkorelasi, meskipun demikian cara ini tidak selalu cukup karena beberapa jenis multikolinieritas mungkin lebih kompleks. Salah satu teknik yang banyak digunakan dalam statistik untuk mendeteksi multikolinieritas adalah VIF (*Variance Inflation Factor*).

Jika nilai VIF melebihi angka 10 maka dapat disimpulkan terdapat multikolinieritas (Gujarati & Porter, 2009). Adapun hipotesis yang digunakan sebagai berikut:

H_0 : $VIF \leq 10$ (Tidak terdapat multikolinieritas)

H_1 : $VIF > 10$ (Terdapat Multikolinieritas)

Statistik uji:

$$VIF_j = \frac{1}{1-R_j^2}; j = 1, 2, \dots, p \quad (2.1)$$

dengan,

R_j^2 = Koefisien determinasi hasil regresi dari variabel bebas X_j dengan variabel bebas lainnya

Keputusan:

Jika $VIF > 10$ maka H_0 ditolak.

Jika setelah dilakukan pengujian ternyata terdapat masalah multikolinieritas antar variabel, cara yang dapat digunakan untuk mengatasi masalah multikolinieritas menurut Ekananda (2015) sebagai berikut:

1. Mengeluarkan variabel yang diperkirakan mempunyai korelasi cukup tinggi dengan variabel lain
2. Mengkombinasikan data *cross-section* dengan *time-series*
3. Penambahan data
4. Transformasi variabel.

Pada penelitian ini cara yang digunakan untuk mengatasi masalah multikolinieritas yakni dengan mengeluarkan variabel yang diperkirakan mempunyai korelasi cukup tinggi dengan variabel lain. Tujuannya adalah untuk mengeluarkan informasi yang *redundant* yang sebenarnya sudah diwakili oleh variabel bebas yang lain.

2.4 Standardisasi Data

Standardisasi data perlu dilakukan ketika data yang digunakan memiliki perbedaan satuan. Apabila data yang digunakan memiliki satuan yang berbeda secara pasti, maka yang harus dilakukan adalah proses standardisasi dengan mengubah data menjadi *Z-score*. Tujuan dari *Z-score* bukan hanya untuk melakukan normalisasi data, tetapi juga bertujuan untuk meningkatkan kinerja dari proses klasifikasi pada data *mining* (Whendasmoro & Joseph, 2022). *Z-score* merupakan suatu perhitungan yang menunjukkan besarnya nilai pada sampel terhadap rata-rata dalam satuan standar deviasi. *Z-score* juga disebut sebagai nilai standar atau nilai baku yang digunakan untuk membantu menghilangkan bias yang

terjadi karena perbedaan satuan pada variabel yang digunakan (Hair dkk., 2010).

Berikut rumus standardisasi data:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}; i = 1, 2, 3 \dots, n; j = 1, 2, \dots, p \quad (2.2)$$

dengan,

z_{ij} = Nilai *Z-score* dari objek ke-*i* pada variabel ke-*j*

x_{ij} = Nilai objek ke-*i* pada variabel ke-*j*

\bar{x}_j = Nilai rata-rata pada variabel ke-*j*

s_j = Nilai standar deviasi pada variabel ke-*j*

2.5 Metode Perhitungan Jarak

Dalam menentukan sebuah *cluster* perlu diketahui tingkat kemiripan dari objek satu dengan yang lain. Kelompok yang terbentuk adalah kelompok yang memuat anggota-anggota yang berdekatan satu sama lain. Ukuran kedekatan diantara masing-masing anggota disebut dengan jarak. Sebelum melakukan pengelompokan objek data maka seharusnya sudah dilakukan pengukuran jarak. Adapun beberapa cara untuk mengukur jarak tergantung dimensi yang digunakan sebagai acuan.

Ukuran kesamaan (*similarity*) dalam analisis *cluster* ditentukan berdasarkan perkiraan sedekat apa objek satu dengan yang lain. Berikut ini adalah ukuran jarak yang dapat digunakan untuk melakukan pengelompokan diantaranya jarak *euclidean*, jarak *manhattan*, jarak *mahalanobis* dan jarak koefisien asosiasi (Gudono, 2011). Pada penelitian ini, ukuran jarak yang akan digunakan penulis yaitu jarak *euclidean*. Menurut Pribadi, Yunus, & Wiguna (2022) kelebihan menggunakan pengukuran *euclidean* yaitu jarak *euclidean* memiliki hasil yang

lebih optimal dibandingkan dengan perhitungan lainnya serta pengukuran *euclidean* lebih umum dilakukan pada analisis *clustering*, sehingga lebih banyak referensi yang dapat dipelajari.

Jarak *euclidean* mengukur kedekatan jarak yang biasa digunakan dalam analisis statistik pada skala data kontinu sepanjang tidak berkorelasi dan memiliki satuan yang sama antar data yang digunakan. Perhitungan dengan metode *euclidean* ialah menghitung akar kuadrat dari jumlah kuadrat selisih dari nilai setiap objek.

Rumus jarak *euclidean* sebagai berikut:

$$d_{i,k} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{kj})^2} \quad (2.3)$$

dengan,

$d_{i,k}$ = Jarak *euclidean* antara objek ke- i dan objek ke- k

x_{ij} = Nilai objek ke- i pada variabel ke- j

x_{kj} = Nilai objek ke- k pada variabel ke- j

p = Banyak variabel yang diamati

2.6 Metode *Clustering*

Metode *clustering* memiliki banyak teknik dalam analisis *cluster* untuk mengelompokkan objek dalam satu *cluster* atau kelompok berdasarkan kesamaan karakteristiknya. Terdapat beberapa metode dalam analisis *cluster* untuk mengklasifikasikan objek dalam suatu kelompok. Pada umumnya, terdapat dua metode yang dapat digunakan dalam pengelompokan yaitu metode hierarki (*hierarchical clustering method*) dan metode non-hierarki (*non-hierarchical clustering method*).

1. Metode Hierarki

Metode hierarki adalah metode analisis kelompok yang menempatkan tingkatan tertentu dalam suatu seperti struktur pohon, karena akumulasi terjadi secara bertahap. Dalam klaster hierarkis, setiap objek adalah klaster yang terpisah, setelah itu dua objek atau klaster terdekat digabungkan menjadi klaster yang lebih kecil. Analisis *cluster* hierarki dapat dilakukan dengan dua jenis pendekatan yaitu penggabungan (*agglomerative*) dan pemisahan (*divisive*) (Johnson & Wichern, 2007).

a. Penggabungan (*agglomerative hierarchical clustering*)

Metode pengelompokan hierarki *agglomerative* adalah setiap objek dianggap menjadi satu *cluster* kemudian antar *cluster* yang memiliki jarak berdekatan bergabung menjadi satu *cluster* besar. Hal ini menghasilkan sebuah kumpulan hierarki data yang dapat digunakan untuk mengelompokkan kumpulan data dan dari hierarki ini kelompok-kelompok tertentu dapat dipilih (Supranto, 2004). Pada metode hierarki *agglomerative* terdapat beberapa teknik diantaranya: *single linkage*, *average linkage*, *complete linkage* dan *ward*.

Pada penelitian ini metode *agglomerative hierarchical clustering* yang digunakan adalah *complete linkage*, dikarenakan beberapa kelebihan yang dimiliki di antaranya menurut Ros & Guillaume (2019) *complete linkage* mempunyai keunggulan dalam *clustering* yaitu kelompok yang stabil dan menghasilkan kelompok yang terlokalisasi dengan baik sedangkan *single linkage* memiliki kriteria yang sangat kuat dan memungkinkan untuk menangani berbagai bentuk dan densitas, namun sangat sensitif terhadap

noise. Selain itu, pada penelitian Dabhi dkk. (2016) *complete linkage* tidak terlalu dipengaruhi oleh *outliers* dan mampu menangani kumpulan data dalam jumlah besar. Oleh karena itu, *complete linkage* bekerja dengan baik dalam memisahkan *cluster* jika terdapat *noise* antar *cluster*.

Complete linkage merupakan teknik pengelompokan yang diukur berdasarkan jarak terjauh antar objek. Teknik ini juga dikenal sebagai pautan lengkap karena semua objek pengamatan dalam suatu *cluster* terhubung ketika jaraknya paling jauh atau paling tidak mirip. Langkah pertama yang dilakukan untuk menghitung menggunakan *complete linkage* adalah memuat elemen minimum dalam matriks di $D = \{d_{i,k}\}$ dan mengelompokkan objek-objek yang sesuai misalkan U dan V , dikelompokkan untuk memperoleh *cluster* (UV).

Langkah berikutnya adalah menghitung jarak antara *cluster* (UV) dengan *cluster* lainnya, misal W , maka akan mendapatkan rumus berikut (Johnson & Wichern, 2007):

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (2.4)$$

dengan,

$$d_{(UV)W} = \text{Jarak antara } cluster (UV) \text{ dan } W$$

$$d_{UW} = \text{Jarak antara } cluster U \text{ dan } W$$

$$d_{VW} = \text{Jarak antara } cluster V \text{ dan } W$$

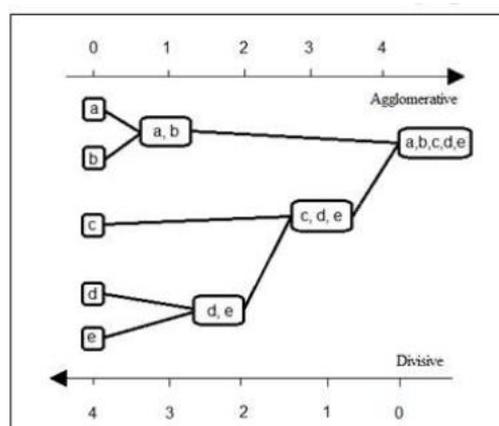
b. Pemisahan (*divisive*)

Metode hierarki *divisive* merupakan kebalikan dari definisi metode *agglomerative*, yaitu kelompok pertama dibagi menjadi dua subkelompok sedemikian rupa, sehingga objek dalam satu subkelompok jauh dari objek

di subkelompok lainnya (Johnson & Wichern, 2007). Singkatnya, metode pemisahan adalah metode hierarki di mana semua objek pengamatan dimulai dari satu kelompok dan dibagi lagi pertama menjadi dua kelompok, kemudian menjadi tiga kelompok, dan seterusnya, sampai semua diberikan subkelompok, yang masing-masing menjadi satu kelompok dengan satu anggota.

Pada dasarnya semua metode *clustering* mempunyai kelemahan dan kelebihan masing-masing, namun secara garis besar analisis *cluster* memiliki kelebihan yaitu dapat mengelompokkan data dalam jumlah besar dan variabel yang banyak, sedangkan kelemahannya adalah perhitungannya lambat, sensitif terhadap pencilon dan sebagainya (Govender & Sivakumar, 2020).

Pengelompokan dengan metode hierarki *agglomerative* dan *divisive* dapat dilihat dengan grafik pohon pada Gambar 2.2 yang biasa disebut dengan *Dendrogram*. *Dendrogram* ini memberikan gambaran *divisive* dan *agglomerative* pada setiap metode analisis.



Gambar 2.2 Contoh *Dendrogram* Metode Hierarki
(Sumber: Kauffman dan Rousseeuw (1990))

2. Metode Non-Hierarki

Berbeda dengan metode hierarki, metode non-hierarki diawali dengan menentukan jumlah kelompok atau *cluster* yang diinginkan. Metode pengelompokan non-hierarki dirancang untuk mengelompokkan item daripada variabel ke dalam kumpulan *cluster*. Jumlah *cluster* dapat ditentukan terlebih dahulu atau ditentukan sebagai bagian dari prosedur *clustering*. *Non-hierarchical method's* biasa dikenal sebagai metode *k-means cluster*. Metode *k-means* bekerja dengan membagi data ke dalam *cluster* yang telah ditentukan oleh penulis kemudian secara iteratif mengamati kembali ke dalam *cluster* sampai beberapa kriteria terpenuhi (Hair dkk., 2010). Sehingga dapat disimpulkan bahwa metode pengelompokan non-hierarki bekerja dengan mengelompokkan objek-objek yang mempunyai karakter yang serupa menjadi satu *cluster* dan mengelompokkan objek-objek yang tidak serupa ke dalam *cluster* yang lain.

2.7 Validasi Analisis *Cluster* berdasarkan *Multiscale Bootstrap*

Setelah memperoleh hasil *clustering*, langkah selanjutnya adalah melakukan perhitungan validitas setiap *cluster*. Menurut Kapita & Abdullah (2020), validasi *cluster* adalah tahapan dalam analisis *cluster* yang berguna untuk memberikan informasi penilaian hasil dari analisis dengan menggunakan pendekatan jumlah atau kuantitatif dan sesuai dengan kondisi yang sebenarnya atau objektif. Salah satu pengukuran yang dapat digunakan untuk memeriksa validitas hasil *clustering* adalah *multiscale bootstrap*. Teknik kerjanya yaitu mengumpulkan sampel yang telah di *bootstrap* kemudian memilih satu diantara perolehan

resampling bootstrap yang dianggap telah mewakili data asli (Effron & Tibshirani, 1993).

Metode ini melakukan *resampling* data dengan membuat sampel ulang yang diamati dengan replikasi B kali. Prinsipnya adalah membuat sampel acak berukuran n dengan mengambil pengamatan dari populasi dengan pengembalian. Dengan kata lain, setiap pengamatan memiliki kesempatan yang sama untuk dipilih dalam setiap sampel ulang.

Adapun dua poin penting dalam menentukan besaran B (Effron & Tibshirani, 1993):

1. Menggunakan replikasi *bootstrap* sedikit, misalkan $B= 25$ dianggap cukup informatif. $B= 50$ sudah cukup untuk memaparkan estimasi simpangan baku dengan baik. Pada analisis *cluster* jumlah replikasi yang dibutuhkan adalah $B= 100$ sampai $B= 500$.
2. Penggunaan $B= 200$ lebih, jarang digunakan untuk memperkirakan standar deviasi, akan tetapi penggunaan B yang semakin besar diperlukan untuk membangun interval kepercayaan *bootstrap*.

Pada penelitian ini besaran B yang digunakan adalah $B = 500$, untuk hasil yang baik dan estimasi waktu replikasi yang tidak terlalu lama. Menurut Shimodaira & Suzuki (2006) mengatakan bahwa $B = 100$ terlalu kecil untuk penggunaan yang sebenarnya, $B = 1000$ atau lebih akan lebih bagus dan berguna untuk diagnosis, namun estimasi waktu yang dibutuhkan juga lama.

Secara umum, hasil dari penggabungan hierarki untuk n mengandung $n - 1$ tahap. Namun, itu masih belum jelas seberapa kuat sebuah *cluster* didukung oleh data. Oleh karena itu, penting untuk mengetahui keakuratan *cluster* tersebut,

dimana akurasi berarti tingkat keyakinan keberadaan *cluster*. Pengukuran akurasi dari sebuah *cluster* menggunakan nilai AU *p-value*, dengan rentang nilai 0 hingga 1. Pada umumnya, konsep *p-value* digunakan untuk menguji hipotesis statistik, dalam konsep ini, jika *p-value* dari suatu *cluster* kurang dari α (tingkat signifikansi yang telah ditentukan), maka *cluster* tersebut ditolak pada tingkat signifikansi α (Suzuki & Shimodaira, 2004).

Multiscale bootstrap adalah metode yang menghitung *p-value* dari hipotesis dengan cara melakukan *resampling* terhadap data. Jadi, *multiscale bootstrap* adalah metode yang digunakan untuk mengukur ketidakpastian dalam *clustering* hierarki dengan melakukan repetisi *bootstrap* pada subset data pada skala yang berbeda (Suzuki & Shimodaira, 2004). Tingkat kepercayaan (*confidence level*) yang digunakan yaitu 95% dengan tingkat signifikansi (α) = 0,05.

Pada pendekatan *multiscale bootstrap* terdapat nilai *bootstrap probability* (BP) dan nilai *approximately unbiased* (AU). Nilai *approximately Unbiased* (AU) memiliki keunggulan dalam hal ketidakbiasan dibandingkan dengan nilai *bootstrap probability* (BP). Suatu *cluster* dengan nilai AU *p-value* $\geq 0,95$ maka dianggap valid, apabila suatu *cluster* dinyatakan valid artinya *cluster* tersebut telah akurat dan dapat menggambarkan populasi yang sebenarnya (Ramly dkk., 2022). Berikut merupakan gambaran dari *hierarchical clustering multiscale bootstrap* (Suzuki & Shimodaira, dalam Laksmi dkk., 2023):

1. Data \mathbf{X} adalah data sampel dengan ukuran n yang terdiri dari $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ dimana \mathbf{X}_i merupakan vektor data pengamatan.
2. Sampel data diambil secara acak dan diperoleh data sampel baru yang diberi label sebagai \mathbf{X}^* .

3. Lakukan langkah 2 secara berulang sebanyak B kali sehingga didapatkan himpunan data *bootstrap* dan didefinisikan sebagai $(\mathbf{X}^{*1}, \mathbf{X}^{*2}, \dots, \mathbf{X}^{*B})$ dimana setiap sampel *bootstrap* nya merupakan sampel acak yang saling independen.

Tahapan perhitungan *approximately unbiased* sebagai berikut (Sharuddin & Muda):

1. Memisalkan N sebagai sampel data asli sedangkan N' adalah sampel data *bootstrap* dimana $N' < N, N' = N$ dan $N' > N$.

Menentukan nilai N' dengan cara:

- a. N' pertama yaitu $N' = \frac{N}{2}$
- b. N' kedua yaitu $N' = N' \text{ pertama} + \left(\frac{N}{m}\right), m = \text{Jumlah iterasi}$
- c. dan seterusnya hingga m iterasi.

dengan,

N = Jumlah data pengamatan

N' = Jumlah data pengamatan *bootstrap*

2. Memisalkan $r = \sqrt{N'/N}$ menjadi panjang urutan relatif dari replikasi *bootstrap* dan m menjadi jumlah iterasi *bootstrap*. Kemudian, tentukan konstanta penskalaan r_m untuk $m = 1, 2, \dots, M$ dan jumlah replikasi *bootstrap* B .
3. Membangkitkan B replikasi *bootstrap* dengan panjang urutan $N' = r_m N$ untuk memperoleh himpunan dari replikasi *bootstrap*, $X^{*1}(r_m), X^{*2}(r_m), \dots, X^{*Bm}(r_m)$.
4. Menentukan nilai v dan c dari hasil perhitungan $z(N')$. Nilai $z(N')$ adalah nilai $z = -\Phi^{-1}(\text{BP})$, dengan $\Phi^{-1}(\cdot)$ merupakan invers dari fungsi

distribusi normal baku $\Phi(\cdot)$. Dengan nilai BP dari sebuah *cluster* adalah frekuensi yang muncul dalam replikasi *bootstrap*. Perhitungan nilai v dan c sebagai berikut:

$$z(N') = v\sqrt{\frac{N'}{N}} + c\sqrt{\frac{N}{N'}} \quad (2.5)$$

Menurut Effron 1985, dalam Shimodaira (2004) nilai v dan c adalah nilai-nilai yang terkait dengan aspek geometris dari hipotesis, dimana v adalah jarak bertanda (*signed distance*) dan c adalah kelengkungan (*curvature*) dari batas. v adalah jarak bertanda (*signed distance*) yang menunjukkan seberapa dekat atau jauh hubungan antar *cluster*, v negatif menunjukkan bahwa titik data yang diamati berada di dalam batas *cluster*, menunjukkan terdapat kemiripan atau kedekatan yang signifikan antar objek atau *cluster*. Begitu juga sebaliknya jika v positif menunjukkan bahwa titik data yang diamati berada di luar batas *cluster* artinya terdapat perbedaan yang signifikan antar objek atau *cluster*. Sedangkan c adalah kelengkungan (*curvature*) yang menunjukkan variasi dan kompleksitas hubungan antar *cluster*, c bernilai negatif artinya batas cenderung melengkung ke arah yang berlawanan dengan jarum jam, menunjukkan struktur yang mungkin kurang kompleks pada tingkat hierarki. c bernilai positif artinya batas cenderung melengkung searah dengan jarum jam, menunjukkan kompleksitas atau variasi antar *cluster*.

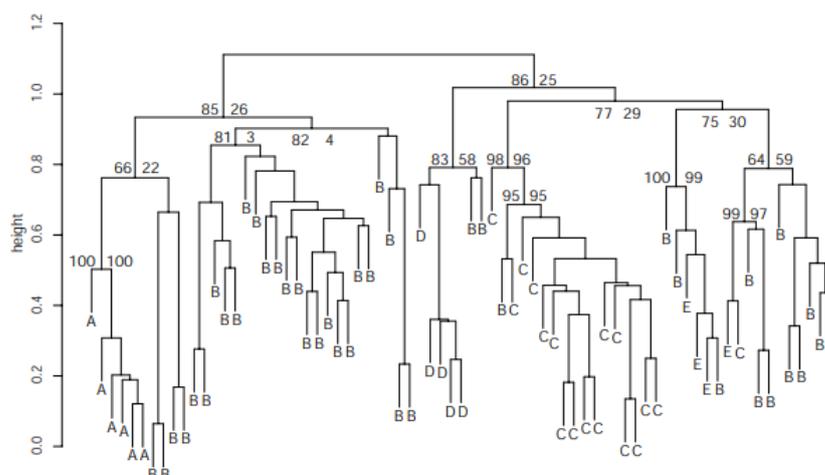
5. Setelah memperoleh nilai v dan c , hitung nilai AU dengan rumus sebagai berikut:

$$AU = \Phi(-v + c) \quad (2.6)$$

dengan,

$\Phi(\cdot)$ = Fungsi distribusi normal baku

Perhitungan nilai AU dan BP bisa juga dilakukan menggunakan *software* statistik R dengan *add-on package* yaitu *Pvclust* yang bertujuan untuk menentukan banyaknya *cluster* optimum serta menghitung AU dan BP untuk masing-masing *cluster* menggunakan teknik *bootstrap resampling* (Suzuki & Shimodaira, 2006).



Gambar 2.3 Contoh *Dendrogram* Metode *Multiscale Bootstrap*
(Sumber: Suzuki Ryota dan Shimodaira Hidetoshi (2004))

Pengelompokan dengan metode *multiscale bootstrap* dapat dilihat pada Gambar 2.3 yaitu grafik pohon yang biasa disebut dengan *dendrogram*. *Dendrogram* ini memberikan gambaran nilai di sisi kiri cabang adalah nilai AU dan pada sisi kanan cabang nilai BP. Label pada daun sebagai contoh (A, B, C, D) adalah objek yang diklasifikasikan.

2.8 Indikator Pendidikan

Salah satu cara untuk mendapatkan pendidikan adalah melalui pendidikan formal. Pemerintah mengaktifkan pendidikan berkualitas dengan mewajibkan sekolah bagi setiap penduduk. Warga negara harus mengikuti program wajib belajar

di bawah tanggung jawab pemerintah pusat dan daerah. Indikator pendidikan adalah variabel atau ukuran yang digunakan untuk mengukur tingkat pendidikan dalam suatu populasi atau kelompok tertentu. Indikator pendidikan memberikan informasi tentang sejauh mana pendidikan telah diterima atau dicapai oleh individu atau kelompok, dan digunakan dalam berbagai konteks untuk memahami tingkat pendidikan, kualitas pendidikan, atau prestasi pendidikan. Beberapa indikator pendidikan adalah sebagai berikut (BPS, 2021).

1. Angka Partisipasi Kasar (APK)

APK adalah persentase siswa pada suatu kelas tertentu dibagi dengan jumlah penduduk pada kelompok usia sekolah tertentu. Semakin tinggi nilai APK, semakin banyak penduduk usia sekolah yang bersekolah di wilayah tersebut. Nilai APK yang mendekati 100 persen atau lebih memberikan pengertian bahwa ada warga yang bersekolah di bawah umur, atau ada yang lebih tua. Hal ini juga menunjukkan bahwa wilayah tersebut mampu menampung penduduk usia sekolah lebih dari target yang sesungguhnya.

APK Rumus perhitungan APK sebagai berikut:

$$APK = \frac{\text{jumlah siswa SMA sederajat}}{\text{jumlah penduduk usia 16-18 tahun}} \times 100$$

2. Angka Partisipasi Sekolah (APS)

Angka partisipasi sekolah adalah persentase proporsi dari semua anak pada kelompok umur tertentu yang masih bersekolah dibagi dengan banyaknya penduduk usia sekolah tertentu. APS ini berguna untuk menunjukkan tingkat partisipasi pendidikan menurut kelompok umur tertentu. Tersedianya sarana dan prasarana pendidikan yang memadai akan mendorong peningkatan partisipasi sekolah diberbagai kelompok usia

sekolah. Tercapainya tujuan pemerintah daerah, khususnya pemerintah Jawa Timur untuk menyelenggarakan pendidikan yang mudah diakses dapat digambarkan melalui besaran APS. Rumus yang digunakan yaitu:

$$APS = \frac{\text{jumlah siswa SMA sederajat yang masih bersekolah}}{\text{jumlah penduduk usia 16 - 18 tahun}} \times 100$$

3. Angka Partisipasi Murni (APM)

APM merupakan persentase jumlah siswa pada kelompok usia tertentu dibagi dengan jumlah penduduk sekolah pada kelompok umur tersebut. Skor APM sama dengan 100 persen ketika semua anak usia sekolah dapat bersekolah tepat waktu. APM digunakan sebagai indikator untuk mengukur proporsi anak yang bersekolah tepat waktu. APM dapat menunjukkan berapa banyak penduduk usia sekolah yang dapat memanfaatkan fasilitas pendidikan sesuai dengan usia pada jenjang pendidikannya. Mengikuti pendidikan dengan usia yang sesuai merupakan langkah persiapan dalam upaya peningkatan mutu pendidikan. Rumus perhitungan APM sebagai berikut:

$$APM = \frac{\text{jumlah siswa SMA sederajat usia 16-18 tahun}}{\text{jumlah penduduk usia 16-18 tahun}} \times 100$$

4. Pendidikan SMA Sederajat yang ditamatkan

Pendidikan tertinggi yang ditamatkan yaitu persentase penduduk SMA sederajat menurut ijazah atau STTB tertinggi yang dimiliki. Salah satu indikator dalam melihat kualitas sumber daya manusianya yaitu dengan melihat pendidikan yang ditamatkan. Pengetahuan seseorang pada umumnya akan meningkat seiring dengan bertambah tingginya tingkat pendidikan yang ditempuh. Karakteristik wilayah berpengaruh pada tingkat pendidikan tertinggi yang ditamatkan, wilayah perkotaan memiliki fasilitas

dan kemudahan akses sarana prasarana yang lebih memadai dan menjadi daya Tarik penduduk di luar wilayah tersebut. Selain itu, dapat digunakan sebagai bahan dalam menganalisis pasar tenaga kerja.

5. Rata-rata Lama Sekolah (RLS)

Rata-rata lama pendidikan merupakan rata-rata lama bersekolah penduduk pada semua jenjang pendidikan formal. Rata-rata lama sekolah dapat digunakan sebagai cara untuk mengidentifikasi kualitas penduduk dalam hal pendidikan formal. Semakin lama rata-rata tahun bersekolah, semakin tinggi kualifikasi pendidikannya. Kemudahan akses pendidikan juga mempengaruhi tingkat rata-rata lama sekolah.

6. Angka Melek Huruf (AMH)

AMH adalah persentase penduduk usia 10 tahun ke atas yang dapat membaca dan menulis dibagi dengan jumlah penduduk usia 10 tahun ke atas. Angka ini menunjukkan capaian daerah terhadap indikator dasar, karena literasi merupakan basis terpenting untuk memperluas pengetahuan.

2.9 Pendidikan dalam Alqur'an

Pendidikan merupakan hal yang penting dan berkesinambungan dengan permasalahan zaman. Hal tersebut dijelaskan dalam UUD 1945 Pasal 28 C ayat 1 mengenai setiap orang berhak mengembangkan diri melalui pemenuhan dasarnya, berhak mendapat pendidikan dan memperoleh manfaat dari ilmu pengetahuan dan teknologi, seni dan budaya, demi meningkatkan kualitas hidupnya dan demi kesejahteraan umat manusia. Maka dari itu, menjadikan manusia yang berkualitas dengan beragam metode efektif dan dinamis serta dapat mengembangkan segala

potensi manusia adalah salah satu tujuan adanya pendidikan. Pendidikan dalam Alqur'an juga dijelaskan secara tersirat pada ayat 1-5 dari surah Al-'Alaq:

﴿ اِقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ (۱) خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ (۲) اِقْرَأْ وَرَبُّكَ الْأَكْرَمُ (۳) الَّذِي عَلَّمَ بِالْقَلَمِ (۴) عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ (۵) ﴾

“(1) Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan, (2) Dia telah menciptakan manusia dari segumpal darah, (3) Bacalah, dan Tuhanmulah Yang Mahamulia, (4) Yang mengajar (manusia) dengan pena, (5) Dia mengajarkan manusia apa yang tidak diketahuinya” (Departemen Agama RI, 2006).

Menurut Quraish Shihab dalam tafsir Al-Mishbah, kata (اِقْرَأْ) terambil dari kata kerja (قَرَأَ) yang pada awalnya berarti menghimpun. Maknanya, ketika seseorang merangkai huruf atau kata kemudian diucapkan rangkaian tersebut artinya seseorang tersebut telah menghimpunnya yakni membacanya. Huruf (ب) pada kata (يَاسْمُ) dipahami dengan penyertaan atau *mulabasah* sehingga dengan demikian ayat tersebut berarti “bacalah disertai dengan nama Tuhanmu”. Kata (رَبِّ) seakar dengan kata (تَرْبِيَةٌ) yang berarti pendidikan. Kata tersebut mengacu pada arti arti pengembangan, peningkatan, ketinggian, kelebihan serta perbaikan.

Manusia sebagai makhluk pertama yang disebut Allah dalam Alqur'an melalui wahyu pertama. Bukan saja karena ia diciptakan dalam bentuk yang sebaik-baiknya atau karena segala sesuatu dalam alam raya ini diciptakan Allah demi kepentingannya, namun juga karena ditujukan kepada manusia untuk menjadi pelita kehidupannya. Salah satu cara yang harus dilakukan manusia adalah dengan menghayati, mempelajari petunjuk-petunjuk Allah dengan menjabarkan proses kejadiannya.

Pada ayat ketiga kata perintah membaca diulangi dengan maksud Allah menjanjikan bahwa pada saat seseorang membaca dengan ikhlas karena Allah,

maka Allah akan menganugerahkan kepadanya ilmu pengetahuan, pemahaman-pemahaman, wawasan-wawasan baru walaupun yang dibacanya itu-itu juga. Apa yang dijanjikan terbukti sangat jelas. Kegiatan “membaca” ayat Alqur’an menimbulkan penafsiran-penafsiran baru atau pengembangan dari pendapat-pendapat yang telah ada. Demikian juga, kegiatan “membaca” alam raya ini telah menimbulkan penemuan-penemuan baru yang membuka rahasia-rahasia alam, walaupun objek bacaannya itu-itu juga. Ayat Alqur’an yang dibaca orang terdahulu adalah sama tidak berbeda, namun pemahaman yang diperoleh akan terus berbeda dan terus berkembang.

Pada ayat keempat dan kelima, Dia Yang Maha Pemurah itu Yang Mengajar manusia dengan pena yakni dengan sarana dan usaha manusia masing-masing, dan Dia juga mengajar manusia tanpa alat dan usaha mereka apa yang belum diketahuinya. Pada awal surah Al-‘Alaq telah disebutkan bahwa Allah sebagai Yang Maha Kuasa, Maha Mengetahui dan Maha Pemurah. Pengetahuan-Nya meliputi segala hal. Sedangkan kemurahan-Nya tidak terbatas, sehingga Allah kuasa dan berkenan untuk mengajar manusia dengan atau tanpa pena (Shihab, 2009).

2.10 Kajian Penelitian dengan Teori Pendukung

Penelitian ini disusun berdasarkan beberapa teori pendukung terkait HCMB. Topik yang dibahas adalah HCMB pada indikator pendidikan tingkat SMA sederajat di Provinsi Jawa Timur. Terdapat beberapa penelitian yang menggunakan metode HCMB. Diantaranya penelitian Ramadhani & Purnamasari (2018) dengan mengaplikasikan *complete linkage* dan *multiscale bootstrap* pada kasus kemiskinan di Kalimantan Timur 2016. Penelitian tersebut menerapkan perbandingan

pengelompokan dengan teknik *complete linkage* dan *multiscale bootstrap*. Hasil yang diperoleh yaitu terdapat empat *cluster* pada kedua metode. Sedangkan pada penelitian Anuraga (2015) dengan topik penelitian HCMB untuk pengelompokan kemiskinan di Jawa Timur. Hasil penelitian yang diperoleh yaitu terdapat lima kelompok yang terbentuk dengan kemiripan satu dengan yang lain dan nilai signifikansi nilai $AU \geq 0,95$. Serta memiliki kestabilan pada saat $B = 500$.

Penggunaan metode hierarki dengan pengukuran jarak hanya memberikan satu solusi dalam penyelesaiannya, yaitu didasarkan pada ukuran kemiripan pada metode jarak yang digunakan. Sehingga penelitian ini menggunakan *multiscale bootstrap* untuk memberikan ukuran ketidakpastian dalam teknik klasik pengelompokan hierarki. Karena itu, pada penelitian ini digunakan metode HCMB untuk mengetahui apakah metode tersebut juga dapat bekerja dengan baik jika digunakan pada pengelompokan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan. Pada penelitian perlu dilakukan uji multikolinieritas untuk memperlihatkan adanya korelasi antar dua atau lebih variabel, selanjutnya perlu dilakukan standardisasi data untuk meningkatkan kinerja dari proses klasifikasi pada data *mining*. Kemudian menentukan ukuran jarak kedekatan antar objek dengan menggunakan jarak *euclidean*. Setelah menentukan jarak kedekatan selanjutnya yaitu menganalisis *cluster* terhadap indikator pendidikan dengan menggunakan teknik *complete linkage* dan melakukan uji validasi menggunakan pendekatan *multiscale bootstrap*. Setelah mendapatkan hasil pengelompokan yang signifikan atau valid kemudian dilakukan interpretasi dan mengambil kesimpulan dari hasil penelitian yang telah dilakukan.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang digunakan yaitu metode deskriptif kuantitatif. Metode deskriptif merupakan pengumpulan informasi data guna menguji ataupun menanggapi pertanyaan. Penelitian kuantitatif merupakan penelitian yang didasarkan pada pengumpulan serta analisis data berbentuk angka yang kemudian akan dianalisis sesuai dengan kebutuhan penelitian. Penelitian kuantitatif menekankan analisisnya pada data-data numerik yang diolah dengan tata cara statistik.

3.2 Data dan Sumber Data

Data yang digunakan pada penelitian ini berbentuk data sekunder. Data yang digunakan berupa data indikator pendidikan di Provinsi Jawa Timur yang diperoleh dari website BPS Jawa Timur tahun 2021 dengan objek pengamatan sebanyak 38 Kabupaten/Kota di Jawa Timur. Data tersebut terdiri dari 6 Indikator diantaranya APK, APS, APM, Pendidikan tingkat SMA yang ditamatkan, RLS, dan AMH.

3.3 Instrumen Penelitian

Instrumen penelitian yang digunakan untuk penelitian yaitu *software* R-studio.

3.4 Variabel Penelitian

Berdasarkan penjelasan mengenai indikator pendidikan pada subbab 2.8, variabel penelitian yang digunakan terdiri dari 6 variabel yang disajikan dalam Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Indikator	Variabel	Satuan
X_1	Angka Partisipasi Kasar (APK)	Persen
X_2	Angka Partisipasi Sekolah (APS)	Persen
X_3	Angka Partisipasi Murni (APM)	Persen
X_4	Pendidikan SMA sederajat yang ditamatkan	Persen
X_5	Rata-rata Lama Sekolah (RLS)	Tahun
X_6	Angka Melek Huruf (AMH)	Persen

3.5 Tahap Analisis Data

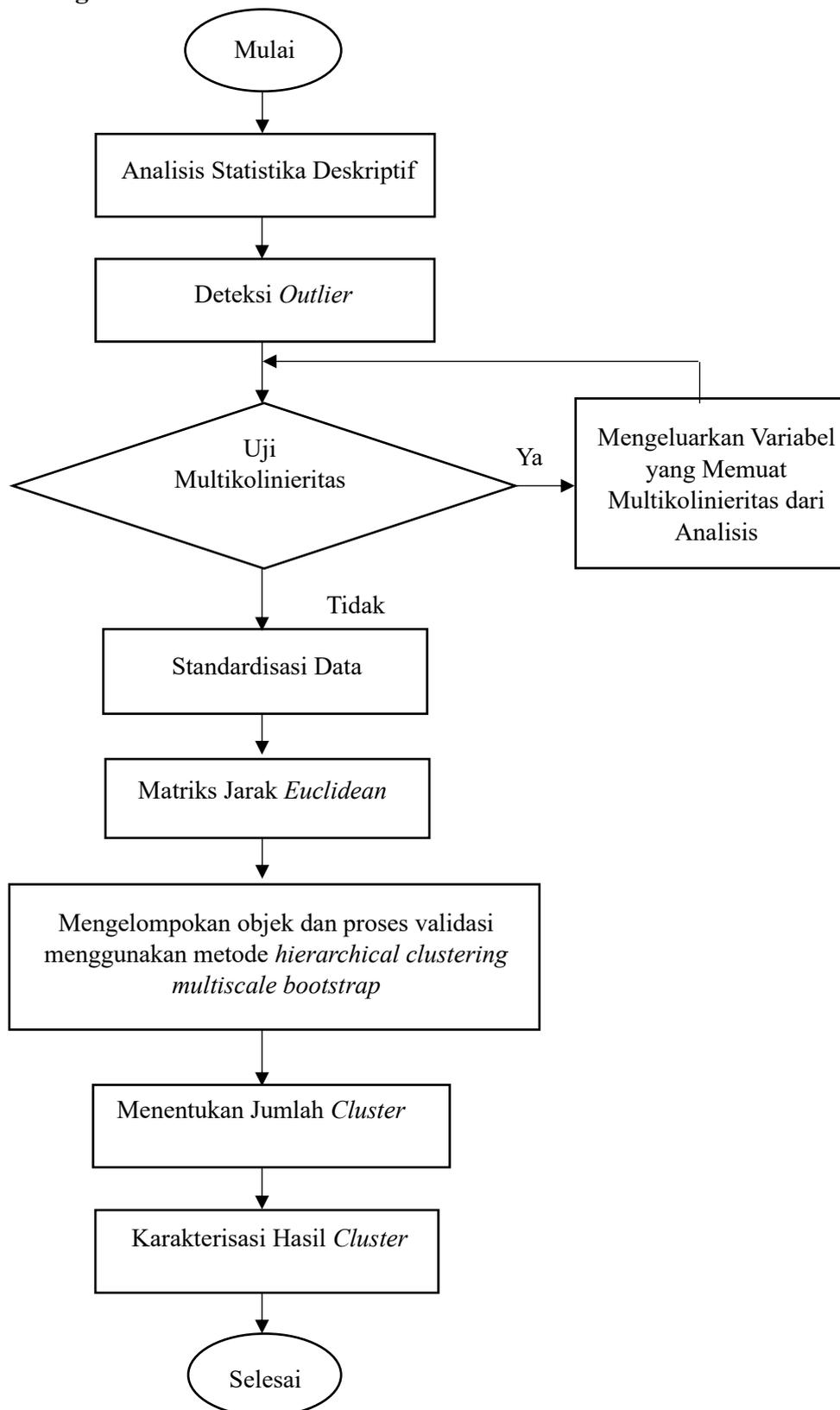
Adapun tahapan dalam penelitian yang menggunakan metode *hierarchical clustering multiscale bootstrap* sebagai berikut.

1. Melakukan analisis statistik deskriptif data pada masing-masing variabel yang digunakan untuk memperlihatkan gambaran karakteristik secara umum pada Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan.
2. Melakukan deteksi *outlier* untuk mengetahui apakah terdapat data yang memiliki perbedaan signifikan antar data lainnya.
3. Sebelum melakukan pengelompokan, hal pertama yang harus dilakukan adalah uji asumsi. Uji asumsi yang harus terpenuhi yaitu uji multikolinieritas menggunakan persamaan (2.1). Hal ini dilakukan untuk mengetahui apakah terdapat hubungan antar variabel atau tidak.
4. Melakukan standarisasi data menggunakan persamaan (2.2).

5. Membuat matriks jarak antar objek menggunakan perhitungan jarak *euclidean* berdasarkan persamaan (2.3).
6. Setelah mendapatkan hasil matriks jarak, selanjutnya melakukan pengelompokan menggunakan metode *hierarchical clustering multiscale bootstrap*.
 - a. Langkah pertama, mencari matriks jarak terkecil untuk pasangan *cluster* terdekat, misalkan pasangan *cluster* paling mirip objek i dan k dengan $D = \{d_{i,k}\}$. Sehingga i dan k dipilih.
 - b. Selanjutnya, menggabungkan *cluster* yang telah dipilih menjadi *cluster* baru kemudian melakukan perhitungan menggunakan rumus (2.4). Setelah perhitungan berakhir, selanjutnya memperbarui masukan dalam matriks jarak dengan cara menghapus baris dan kolom sesuai dengan *cluster* yang diperoleh dengan memberikan nilai jarak antar *cluster* baru dan semua sisa *cluster*, sehingga akan diperoleh matriks jarak baru.
 - c. Ulangi langkah a dan b sebanyak $(n-1)$ kali. Sehingga, semua objek akan berada dalam *cluster* tunggal sampai berakhirnya algoritma.
 - d. Setelah proses penggabungan berakhir tahapan berikutnya adalah memisalkan N sebagai sampel data asli sedangkan N' adalah sampel data *bootstrap* dimana $N' < N, N' = N$ dan $N' > N$.
 - e. Memisalkan $r = \sqrt{N'/N}$ menjadi panjang urutan relatif dari replikasi *bootstrap* dan m menjadi jumlah iterasi *bootstrap*. Kemudian, tentukan konstanta penskalaan r_m untuk $m = 1, 2, \dots, M$ dan jumlah replikasi *bootstrap* B .

- f. Membangkitkan B replikasi *bootstrap* dengan panjang urutan $N' = r_m N$ untuk memperoleh himpunan dari replikasi *bootstrap*, $X^{*1}(r_m), X^{*2}(r_m), \dots, X^{*Bm}(r_m)$.
 - g. Menghitung nilai z yang diamati dengan menyesuaikan kurva teoritis menggunakan persamaan (2.5) untuk mengestimasi nilai v dan c .
 - h. Selanjutnya menghitung nilai AU menggunakan persamaan (2.6).
7. Menentukan banyaknya *cluster* menggunakan nilai selisih koefisien jarak terbesar pada langkah penggabungan.
 8. Karakterisasi hasil anggota pengelompokan pada masing-masing *cluster*.

3.6 Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Tahap Analisis Data

BAB IV

HASIL DAN PEMBAHASAN

4.1 Statistika Deskriptif Karakteristik Indikator Pendidikan SMA Sederajat Kabupaten/Kota di Provinsi Jawa Timur Tahun 2021

Statistika deskriptif digunakan untuk menganalisis data dengan cara menjelaskan data yang telah terkumpul sebagaimana adanya tanpa membuat kesimpulan yang umum atau generalisasi.

Tabel 4.1 Statistik Deskriptif Data Indikator Pendidikan Tingkat SMA Sederajat di Provinsi Jawa Timur

Indikator	N	Minimum	Maksimum	Rata-Rata	Standar Deviasi
Angka Partisipasi Kasar	38	57,00	125,90	87,43	14,71
Angka Partisipasi Sekolah	38	51,28	96,82	76,25	10,32
Angka Partisipasi Murni	38	38,45	85,02	65,14	11,80
Pendidikan SMA yang ditamatkan	38	11,90	43,05	25,15	7,79
Rata-rata Lama Sekolah	38	5,91	11,93	8,59	1,52
Angka Melek Huruf	38	81,69	99,12	93,31	4,71

Tabel 4.1 menunjukkan bahwa sebanyak (*N*) 38 Kabupaten/Kota ada di Jawa Timur. Nilai minimum sebesar 57% dan nilai maksimum 125,9% pada variabel angka partisipasi kasar (APK) menunjukkan bahwa ada penduduk Kabupaten/Kota di Jawa Timur yang bersekolah tingkat SMA sederajat di bawah umur, atau ada warga yang lebih tua. Nilai angka partisipasi kasar pada Kabupaten/Kota memiliki rata-rata sebesar 87,43%. Hal ini menunjukkan bahwa tingkat partisipasi kasar di Jawa Timur tahun 2021 sudah tergolong tinggi, namun

belum mencapai angka partisipasi kasar yang telah direncanakan oleh pemerintah yaitu 88,39%.

Hasil uji statistik angka partisipasi sekolah (APS) memiliki nilai maksimum sebesar 96,82% dan nilai minimum sebesar 51,28%. Nilai rata-rata sebesar 76,25% dan standar deviasi lebih rendah dari pada rata-rata menunjukkan bahwa sebaran data variabel angka partisipasi sekolah pada Kabupaten/Kota bersifat homogen atau sama.

Hasil uji statistik variabel angka partisipasi murni (APM) tingkat SMA sederajat menghasilkan nilai maksimum sebesar 85,02% dan nilai minimum 38,45%. Rata-rata sebesar 65,14%. Nilai APM yang baik atau ideal adalah 100%. Semakin mendekati 100% semakin baik pula tingkat partisipasi murninya. Memiliki nilai rata-rata lebih besar dari nilai standar deviasi mengindikasikan bahwa sebaran data variabel angka partisipasi murni pada Kabupaten/Kota bersifat homogen atau sama.

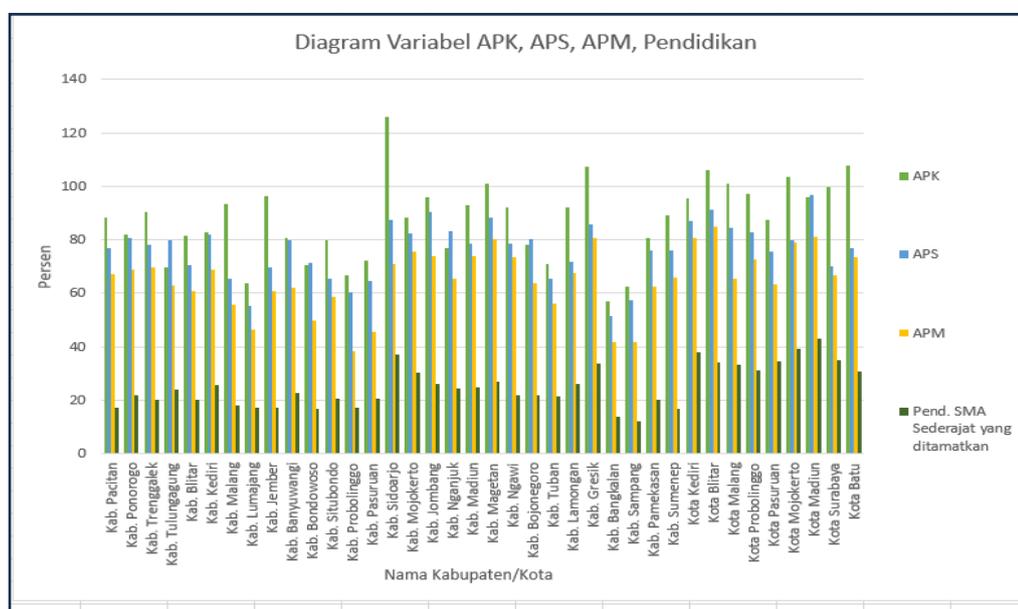
Variabel pendidikan SMA sederajat yang ditamatkan memiliki nilai maksimum sebesar 43,05% dan nilai minimum 11,90%. Hal ini menunjukkan bahwa penduduk yang menyelesaikan pendidikan SMA sederajat hingga lulus dan memiliki ijazah masih perlu ditingkatkan dikarenakan banyak penduduk yang tidak menyelesaikan pendidikan SMA hingga lulus, terbukti dengan selisih jumlah angka partisipasi dengan jumlah penduduk yang menyelesaikan pendidikan hingga mendapatkan ijazah cukup tinggi. Nilai rata-rata 25,15% dan standar deviasi 7,79% menunjukkan sebaran data untuk variabel pada Kabupaten/Kota hampir sama.

Hasil uji statistik variabel rata-rata lama sekolah (RLS) memiliki nilai maksimum sebesar 11,93 tahun dan nilai minimum 5,91 tahun. Hal ini

menunjukkan bahwa jumlah tahun yang digunakan penduduk untuk melaksanakan pendidikan memiliki nilai minimum 6 tahun, maksimum 12 tahun dengan rata-rata 9 tahun. Rata-rata lama sekolah menunjukkan bahwa semakin lama rata-rata tahun bersekolah, semakin tinggi kualifikasi pendidikannya.

Pada variabel angka melek huruf (AMH) memiliki nilai maksimum 99,12% dan nilai minimum 81,69%. Rata-rata angka melek huruf sebesar 93,316% dan standar deviasi 4,754% menunjukkan sebaran data untuk variabel angka melek huruf pada Kabupaten/Kota hampir sama. Semakin tinggi angka melek huruf di suatu daerah menunjukkan capaian daerah tersebut terhadap indikator dasar, karena literasi merupakan basis terpenting untuk memperluas pengetahuan.

Selain dalam bentuk tabel, statistika deskriptif juga dapat digambarkan dengan diagram. Diagram pada variabel APK, APS dan APM yang digunakan dalam penelitian ini ditunjukkan pada Gambar 4.1.

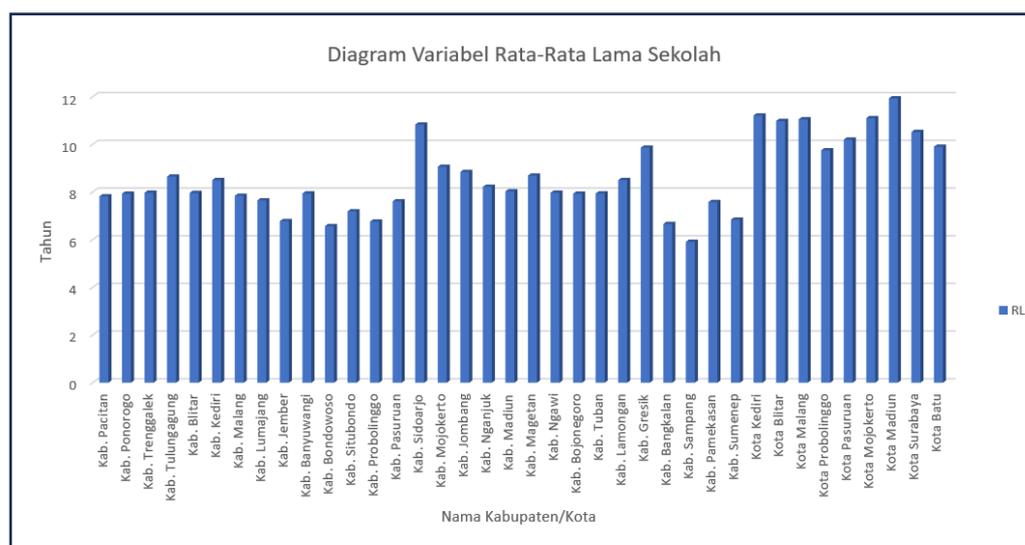


Gambar 4.1 Diagram Variabel APK, APS, APM dan Pend. SMA Sederajat yang Ditamatkan

Berdasarkan Gambar 4.1 menunjukkan bahwa variabel angka partisipasi kasar di Jawa Timur, memiliki angka tertinggi yaitu 125,90% terdapat di Kabupaten Sidoarjo disusul Kota Batu, Kabupaten Gresik, Kota Blitar, Kota Mojokerto, Kota Malang dan Kabupaten Magetan. APK lebih dari 100% bermakna seluruh siswa SMA sederajat mencakup anak-anak diluar batas usia sekolah SMA sederajat 16-18 tahun. Sedangkan APK terendah 57% di wilayah Kabupaten Bangkalan, untuk APS memiliki angka tertinggi yaitu 96,82% terdapat di Kabupaten Madiun dan terendah yaitu 51,28% Kabupaten Bangkalan. Kabupaten Madiun dikatakan cukup baik dalam memberikan akses pendidikan, tercermin dari tingginya APS sebesar 96,82%, menandakan bahwa hampir semua penduduk telah meraih kesempatan untuk mengakses pendidikan.

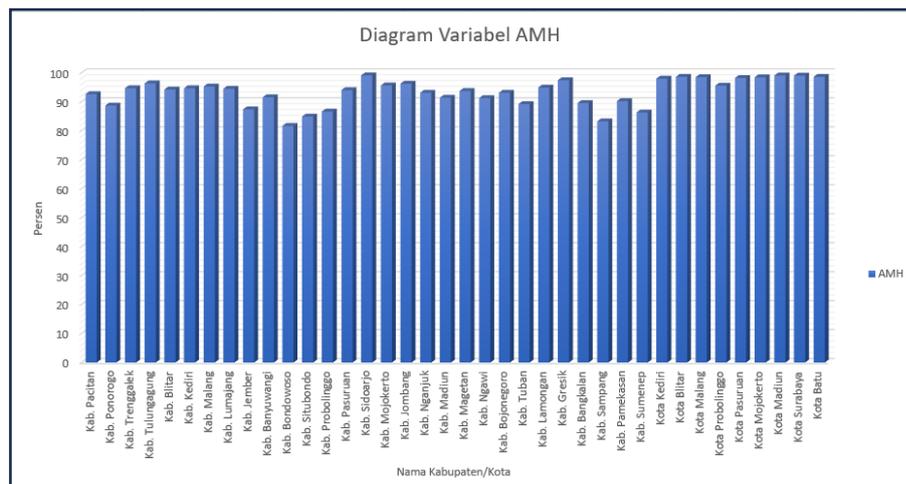
APM di Jawa Timur tertinggi terletak di Kota Blitar dengan semua anak usia sekolah bersekolah sesuai dengan umur sekolah SMA sederajat atau 16-18 tahun sebesar 85,02% dan Kabupaten Probolinggo memiliki nilai terendah sebesar 38,45% di Jawa Timur. Kabupaten Probolinggo memiliki nilai terendah dikarenakan banyak anak yang tidak langsung melanjutkan pendidikan SMA sederajat setelah lulus SMP sederajat. Sebanyak 85,02% penduduk di Kota Blitar memiliki kesadaran untuk bersekolah tepat pada waktunya. Penduduk yang menempuh pendidikan SMA sederajat hingga tamat dan memiliki ijazah tertinggi terletak di Kota Madiun sebesar 43,05% dan terendah terletak di Kabupaten Sampang sebesar 11,90%. Jenjang pendidikan yang ditamatkan seseorang ditandai dengan sertifikat/ijazah. Sebanyak 43,05% penduduk di Kota Madiun telah menyelesaikan pendidikan SMA sederajat hingga lulus dan memperoleh ijazah.

Gambar 4.2 menunjukkan bahwa rata-rata lama sekolah tertinggi yaitu 11,93 tahun terdapat di Kota Madiun dan terendah sebesar 5,91 tahun di Kabupaten Sampang. Pada wilayah perkotaan, rata-rata lama sekolah untuk penduduk usia 15 tahun ke atas cukup tinggi, dikarenakan rata-rata lama sekolah di hampir seluruh wilayah perkotaan memiliki rata-rata 10 tahun atau lebih atau setara dengan kelas 1 SMA sederajat. Hal ini menunjukkan bahwa kemudahan mengakses pendidikan ternyata masih terpusat pada wilayah perkotaan. Hampir semua Kabupaten/Kota di Jawa Timur sudah memiliki rata-rata sekolah di atas 6 tahun kecuali Kabupaten Sampang, hal ini menunjukkan bahwa penduduk di semua Kabupaten/Kota di Jawa Timur rata-rata sudah menyelesaikan wajib belajar sampai SD sederajat.



Gambar 4.2 Diagram Variabel Rata-Rata Lama Sekolah

Gambar 4.3 menunjukkan bahwa variabel angka melek huruf memiliki nilai tertinggi sebesar 99,12% di Kota Sidoarjo dan terendah sebesar 81,69% di Kabupaten Bondowoso. Sebesar 99,12% penduduk di Kota Sidoarjo dan 81,69% penduduk di Kabupaten Bondowoso memiliki kemampuan dasar ilmu pengetahuan yaitu membaca dan menulis.



Gambar 4.3 Diagram Variabel Angka Melek Huruf

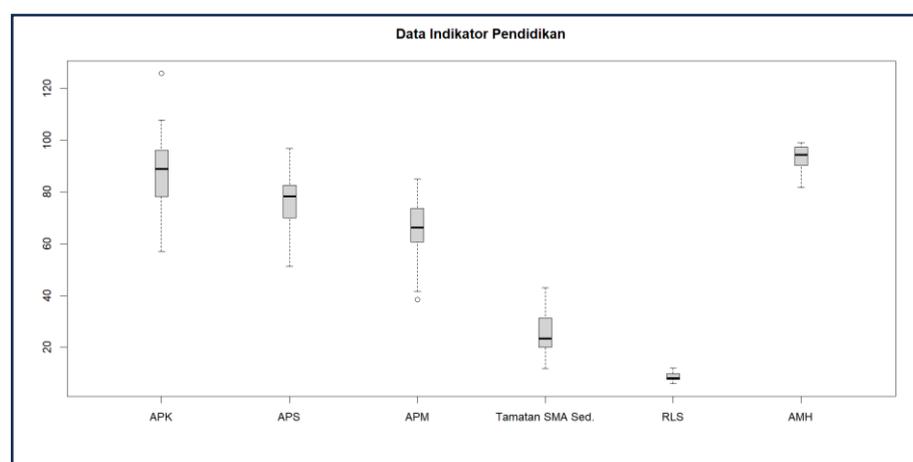
4.2 Deteksi *Outlier*

Sebelum melakukan uji asumsi pada data, terlebih dahulu perlu dilakukan deteksi *outlier*. *Outlier* dideskripsikan sebagai perbedaan data yang signifikan dibanding dengan nilai sekumpulan data yang lain. Hasil visualisasi *boxplot* untuk mendeteksi *outlier* ditunjukkan pada Gambar 4.4.

Pada Gambar 4.4 terlihat bahwa variabel yang terbebas dari *outlier* adalah variabel APS, Pendidikan SMA sederajat yang ditamatkan, RLS dan AMH, ditunjukkan dengan tidak terdapat titik di luar *boxplot*. Sedangkan pada variabel APK dan APM terindikasi *outlier*, ditunjukkan dengan terdapatnya titik di luar *boxplot*. Adanya *outlier* pada variabel APK dikarenakan ada perbedaan data yang signifikan dibanding dengan nilai sekumpulan data yang lain. APK terdeteksi *outlier* pada Kabupaten Sidoarjo memiliki persentase nilai APK yang tinggi di atas 120%.

Sedangkan pada variabel APM juga terlihat ada titik di luar *boxplot*, hal ini menandakan bahwa terdapat *outlier* pada variabel APM. *Outlier* pada variabel APM dikarenakan terdapat objek yang memiliki persentase APM yang terlalu rendah di

bawah 40% yaitu Kabupaten Probolinggo. Skor APM yang baik meningkat hingga 100 persen, dengan semua anak usia sekolah bersekolah tepat waktu berdasarkan usia SMA sederajat. Seperti yang telah dibuktikan pada artikel Badan Pusat Statistik dengan judul “Statistik Pendidikan Provinsi di Jawa Timur Tahun 2021”. Variabel yang terdeteksi *outlier* tetap harus disertakan dalam analisis *cluster*, karena *outlier* mewakili kelompok yang valid dan relevan berdasarkan data BPS dalam artikel “Statistik Pendidikan Provinsi di Jawa Timur Tahun 2021”.



Gambar 4.4 Visualisasi *Boxplot* Data Indikator Pendidikan SMA Sederajat di Provinsi Jawa Timur

4.3 Uji Multikolinieritas

Sebelum melakukan analisis *cluster*, uji asumsi yang harus dipenuhi yaitu uji asumsi non-multikolinieritas dengan tujuan untuk mengetahui ada tidaknya hubungan antar variabel. Apabila data diketahui mengandung multikolinieritas maka data tersebut tidak dapat digunakan untuk proses analisis *cluster*. Maka dari itu, pengujian multikolinieritas perlu dilakukan. Hasil uji asumsi multikolinieritas ditunjukkan dalam Tabel 4.2.

Tabel 4.2 Uji Multikolinieritas Variabel APK, APS, APM, Pend. SMA Sederajat yang Ditamatkan, RLS, AMH

Nilai VIF	APK	APS	APM	Pend. SMA Sederajat yang Ditamatkan	RLS	AMH
APK	-	4,84	5,30	23,12	31,90	4,89
APS	3,06	-	3,24	23,11	32,05	4,81
APM	2,44	2,36	-	22,67	31,77	4,83
Pend. SMA Sederajat yang Ditamatkan	3,09	4,89	6,58	-	5,83	4,15
RLS	3,06	4,86	6,62	4,18	-	2,83
AMH	3,09	4,81	6,62	19,62	18,61	-

Berdasarkan Tabel 4.2 hasil yang diperoleh dalam perhitungan VIF berturut-turut mengindikasikan adanya multikolinieritas dikarenakan nilai VIF > 10. Sehingga dapat disimpulkan bahwa terdapat multikolinieritas antar variabel, maka dari itu perlu dilakukan pengeluan 1 variabel yang memiliki nilai VIF tertinggi yaitu rata-rata lama sekolah. Variabel rata-rata lama sekolah dianggap variabel yang *redundant* karena memiliki pengertian yang serupa dengan variabel pendidikan SMA sederajat yang ditamatkan. Oleh karena itu pengujian multikolinieritas dilakukan kembali dengan 5 variabel yang digunakan yaitu APK, APS, APM, Pendidikan SMA Sederajat yang ditamatkan dan AMH. Hasil pengujian nilai VIF dengan 5 variabel ditunjukkan pada Tabel 4.3.

Tabel 4.3 Uji Multikolinieritas Variabel APK, APS, APM, Pend. SMA Sederajat yang Ditamatkan dan AMH

Nilai VIF	APK	APS	APM	Pend. SMA Sederajat yang Ditamatkan	AMH
APK	-	4,81	5,26	4,03	2,79
APS	3,03	-	3,21	3,97	2,80
APM	2,43	2,36	-	4,16	2,82
Pend. SMA Sederajat yang Ditamatkan	2,94	4,61	6,58	-	1,75
AMH	3,02	4,81	6,60	2,59	-

Setelah dilakukan perhitungan VIF kembali, semua hasil uji $VIF \leq 10$. Sehingga dapat ditarik kesimpulan bahwa data tidak terindikasi adanya multikolinieritas antar variabel sehingga analisis *cluster* dapat dilakukan.

4.4 Standardisasi Data

Standardisasi data disajikan dalam bentuk *z-score* dikarenakan analisis *cluster* hanya menggunakan variabel bebas (independen). Standardisasi dilakukan untuk meningkatkan kinerja dari proses klasifikasi pada data *mining*. Hasil perhitungan standardisasi data dapat dilihat pada lampiran 2. Sebagai contoh, berikut beberapa perhitungan *z-score* pada variabel APK:

$$z_{11} = \frac{x_{11} - \bar{x}_1}{s_1}$$

$$z_{11} = \frac{88,36 - 87,43}{14,7} = 0,06$$

Sehingga nilai *z-score* objek 1 (Kabupaten Pacitan) pada variabel APK adalah 0,06. Selanjutnya menghitung nilai *z-score* dari objek 2 (Kabupaten Ponorogo) pada variabel APK:

$$z_{21} = \frac{x_{21} - \bar{x}_1}{s_1}$$

$$z_{21} = \frac{81,75 - 87,43}{14,7} = -0,39$$

Sehingga nilai *z-score* dari objek 2 pada variabel APK adalah -0,38.

4.5 Matriks Jarak Kedekatan *Euclidean*

Setelah memperoleh nilai standardisasi pada data, selanjutnya adalah menentukan nilai kedekatan antar objek. Pada penelitian ini jarak kedekatan yang

digunakan yaitu perhitungan jarak *euclidean*, rumus menghitung jarak kedekatan *euclidean* berdasarkan persamaan 2.3. Nilai kedekatan antar objek diperoleh dari perhitungan objek ke- i dan objek ke- k , dengan $i, k = 1, 2, 3, \dots, 38, i \neq k$. Sebagai contoh perhitungan antar Kabupaten Pacitan sebagai objek ke-1 dan Kabupaten Ponorogo sebagai objek ke-2, sebagai berikut:

$$d_{12} = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2 + \dots + (x_{15} - x_{25})^2}$$

$$d_{12} = \sqrt{(0,06 - (-0,39))^2 + (0,07 - 0,44)^2 + \dots + ((-0,14) - (-0,98))^2}$$

$$= 1,19$$

Sehingga ukuran jarak pada objek ke-1 (Kabupaten Pacitan) dan objek ke-2 (Kabupaten Ponorogo) adalah 1,19. Selanjutnya menghitung ukuran jarak Kabupaten Pacitan sebagai objek ke-1 dan Kabupaten Trenggalek sebagai objek ke-3, sebagai berikut:

$$d_{13} = \sqrt{(x_{11} - x_{31})^2 + (x_{12} - x_{32})^2 + \dots + (x_{15} - x_{35})^2}$$

$$d_{13} = \sqrt{(0,06 - 0,19)^2 + (0,07 - 0,16)^2 + \dots + ((-0,13) - 0,29)^2} = 0,62$$

Sehingga ukuran jarak pada objek ke-1 (Kabupaten Pacitan) dan objek ke-3 (Kabupaten Trenggalek) adalah 0,62. Selanjutnya menghitung ukuran jarak Kabupaten Ponorogo sebagai objek ke-2 dan Kabupaten Trenggalek sebagai objek ke-3, sebagai berikut:

$$d_{23} = \sqrt{(x_{21} - x_{31})^2 + (x_{22} - x_{32})^2 + \dots + (x_{25} - x_{35})^2}$$

$$d_{23} = \sqrt{((-0,38) - 0,19)^2 + (0,43 - 0,16)^2 + \dots + ((-0,97) - 0,29)^2} = 1,44$$

Sehingga ukuran jarak pada objek ke-2 (Kabupaten Ponorogo) dan objek ke-3 (Kabupaten Trenggalek) adalah 1,44. Diketahui bahwa kedekatan Kabupaten Ponorogo dan Kabupaten Trenggalek lebih tidak mirip daripada dengan jarak

kedekatan Kabupaten Pacitan dan Kabupaten Trenggalek. Hal ini menandakan bahwa jika dua objek mempunyai jarak yang cukup dekat, maka dua objek tersebut memiliki tingkat kemiripan yang tinggi dan begitupun untuk semua objek. Hasil matriks jarak *euclidean* selengkapnya dapat dilihat pada lampiran 3. Berikut adalah ringkasan matriks jarak kedekatan objek yang lainnya:

$$\begin{array}{l}
 d_{euclidean} \\
 d_{(1, k)} \\
 d_{(2, k)} \\
 d_{(3, k)} \\
 d_{(4, k)} \\
 d_{(5, k)} \\
 d_{(6, k)} \\
 \vdots \\
 d_{(38, k)}
 \end{array}
 \begin{array}{cccccccc}
 d_{(i, 1)} & d_{(i, 2)} & d_{(i, 3)} & d_{(i, 4)} & d_{(i, 5)} & d_{(i, 6)} & \cdots & d_{(i, 38)} \\
 \left(\begin{array}{cccccccc}
 0,00 & 1,19 & 0,62 & 1,78 & 1,07 & 1,34 & \cdots & 2,58 \\
 1,19 & 0,00 & 1,44 & 1,91 & 1,70 & 1,37 & \cdots & 3,03 \\
 0,62 & 1,44 & 0,00 & 1,66 & 1,22 & 1,00 & \cdots & 2,04 \\
 1,78 & 1,91 & 1,66 & 0,00 & 1,40 & 1,12 & \cdots & 2,94 \\
 1,07 & 1,70 & 1,22 & 1,40 & 0,00 & 1,49 & \cdots & 2,74 \\
 1,34 & 1,37 & 1,00 & 1,12 & 1,49 & 0,00 & \cdots & 2,11 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 2,58 & 3,03 & 2,04 & 2,94 & 2,74 & 2,11 & \cdots & 0,00
 \end{array} \right)
 \end{array}$$

4.6 Analisis Agglomerative Metode Hierarchical Clustering Multiscale Bootstrap

Setelah melakukan data *pre-processing* dan memperoleh nilai ukuran jarak kedekatan, selanjutnya akan dilakukan pengelompokan hierarki dengan metode HCMB. Pertama yang dilakukan yaitu menentukan ukuran nilai jarak kedekatan *euclidean*, kemudian dibentuk menjadi sebuah matriks jarak. Setelah matriks terbentuk, selanjutnya mencari jarak kelompok yang terkecil dan menggabungkannya menjadi sebuah *cluster*.

Pada teknik ini, jarak antar *cluster* diukur berdasarkan jarak terjauh anggota-anggota mereka. Rumus perhitungan *complete linkage* menggunakan persamaan

2.4. Berdasarkan matriks jarak pada lampiran 3 diketahui jarak paling kecil berada pada objek ke-10 dan objek ke-22 yaitu 0,40. Objek ke-10 dan objek ke-22 digabung menjadi *cluster* 1. Sedangkan untuk mendapatkan *cluster* yang lainnya, diperlukan jarak-jarak antara objek ke-10 dan objek ke-22 serta nilai maksimum dari objek ke- w , dengan $w = 1, 2, 3, 4 \dots 38, w \neq 10, 22$. Berikut adalah contoh perhitungan *complete linkage*:

$$\begin{aligned} d_{(10, 22)1} &= \max(d_{10, 1}, d_{22, 1}) \\ &= \max(1,02; 1,01) \\ &= 1,02 \end{aligned}$$

$$\begin{aligned} d_{(10, 22)2} &= \max(d_{10, 2}, d_{22, 2}) \\ &= \max(0,85; 1,06) \\ &= 1,06 \end{aligned}$$

⋮

$d_{(10, 22)w}$, dengan $w = 1, 2, 3, \dots, 38, w \neq 10, 22$

Selanjutnya menghapus baris-baris dan kolom-kolom matriks yang bersesuaian dengan objek ke-10 dan objek ke-22 serta menambahkan baris dan kolom baru untuk *cluster* 1, sehingga didapatkan matriks jarak baru sebagai berikut:

$$\begin{array}{l} d_{euclidean} \\ d_{(10, 22)} \\ d_{(1, k)} \\ d_{(2, k)} \\ d_{(3, k)} \\ d_{(4, k)} \\ \vdots \\ d_{(38, k)} \end{array} \begin{array}{cccccc} d_{(10, 22)} & d_{(i, 1)} & d_{(i, 2)} & d_{(i, 3)} & d_{(i, 4)} & \dots & d_{(i, 38)} \\ \left(\begin{array}{cccccc} 0 & 1,02 & 1,06 & 1,20 & 1,28 & \dots & 2,79 \\ 1,02 & 0 & 1,19 & 0,62 & 1,78 & \dots & 2,58 \\ 1,06 & 1,19 & 0 & 1,44 & 1,91 & \dots & 3,03 \\ 1,20 & 0,62 & 1,44 & 0 & 1,66 & \dots & 2,04 \\ 1,28 & 1,78 & 1,91 & 1,66 & 0 & \dots & 2,94 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 2,79 & 2,58 & 3,03 & 2,04 & 2,94 & \dots & 0 \end{array} \right) \end{array}$$

Selanjutnya dipilih jarak paling kecil berikutnya berdasarkan matriks baru di atas adalah 0,41. Jarak tersebut terletak pada objek ke-19 dan objek ke-21 yang digabung menjadi *cluster* ke-2 sehingga matriks jarak yang terbentuk sebagai berikut:

$$d_{euclidean} \begin{matrix} d_{(19, 21)} & d_{(10, 22)} & d_{(i, 1)} & d_{(i, 2)} & d_{(i, 3)} & \cdots & d_{(i, 38)} \\ d_{(19, 21)} & \left(\begin{array}{ccccccc} 0 & 1,45 & 1,24 & 1,14 & 1,04 & \cdots & 2,21 \\ 1,45 & 0 & 1,02 & 1,06 & 1,20 & \cdots & 2,79 \\ 1,24 & 1,02 & 0 & 1,19 & 0,62 & \cdots & 2,58 \\ 1,14 & 1,06 & 1,19 & 0 & 1,44 & \cdots & 3,03 \\ 1,04 & 1,20 & 0,62 & 1,44 & 0 & \cdots & 2,04 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 2,21 & 2,79 & 2,58 & 3,03 & 2,04 & \cdots & 0 \end{array} \right) \end{matrix}$$

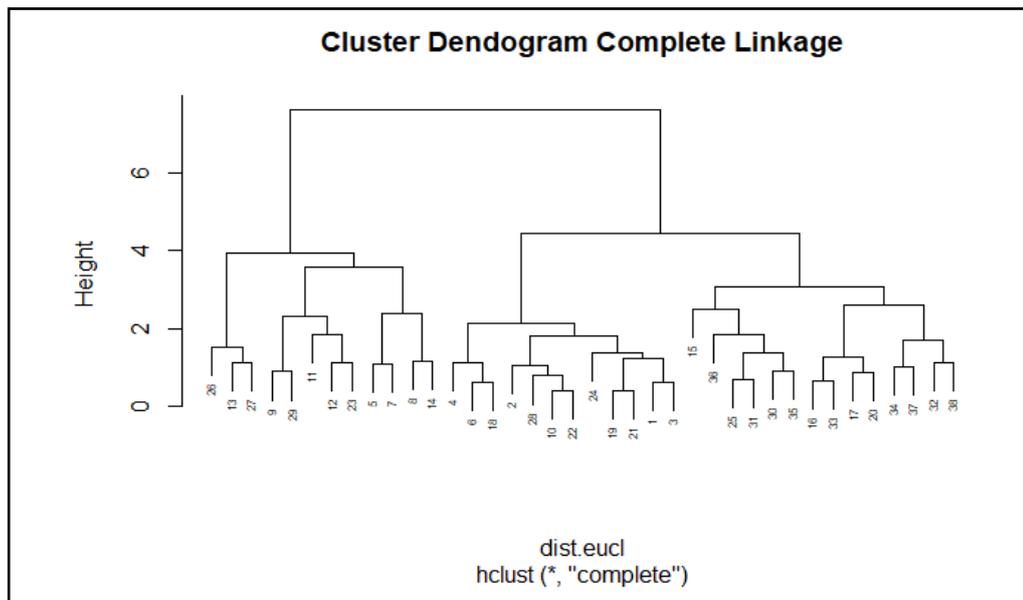
Tahap penggabungan dengan *complete linkage* di atas terus dilakukan hingga matriks terakhir terbentuk. Berikut ini adalah langkah-langkah perhitungan menggunakan penggabungan (*agglomerative*) berdasarkan teknik *complete linkage*:

1. Langkah 1, objek yang memiliki jarak kedekatan minimum digabungkan menjadi *cluster* ke-1, yaitu objek ke-10 dan objek ke-22.
2. Langkah 2, objek ke-19 dan objek ke-21 digabung menjadi *cluster* ke-2.
3. Langkah 3, objek ke-1 dan objek ke-3 digabung menjadi *cluster* ke-3.
4. Langkah 4, objek ke-6 dan objek ke-18 digabung menjadi *cluster* ke-4.
5. Langkah 5, objek ke-16 dan objek ke-33 digabung menjadi *cluster* ke-5.
6. Langkah 6, objek ke-25 dan objek ke-31 digabung menjadi *cluster* ke-6.
7. Langkah 7, objek ke-28 digabungkan dengan *cluster* ke-1, sehingga *cluster* ini terdiri (10, 22, 28).

8. Langkah 8, objek ke-17 dan objek ke-20 digabung menjadi *cluster* ke-7.
9. Langkah 9, objek ke-9 dan objek ke-29 digabung menjadi *cluster* ke-8.
10. Langkah 10, objek ke-30 dan objek ke-35 digabung menjadi *cluster* ke-9.
11. Langkah 11, objek ke-34 dan objek ke-37 digabung menjadi *cluster* ke-10.
12. Langkah 12, objek ke-2 digabung dengan *cluster* ke-1, sehingga *cluster* ini terdiri dari (2, 10, 22, 28).
13. Langkah 13, objek ke-5 dan objek ke-7 digabung menjadi *cluster* ke-11.
14. Langkah 14, objek ke-4 digabungkan dengan *cluster* ke-4, sehingga *cluster* ini terdiri dari (4, 6, 18).
15. Langkah 15, objek ke-13 dan objek ke-27 digabung menjadi *cluster* ke-12.
16. Langkah 16, objek ke-12 dan objek ke-23 digabung menjadi *cluster* ke-13.
17. Langkah 17, objek ke-32 dan objek ke-38 digabung menjadi *cluster* ke-14.
18. Langkah 18, objek ke-8 dan objek ke-14 digabung menjadi *cluster* ke-15.
19. Langkah 19, *cluster* ke-2 digabungkan dengan *cluster* ke-3, sehingga *cluster* ini terdiri dari (1, 3, 19, 21).
20. Langkah 20, *cluster* ke-5 digabungkan dengan *cluster* ke-7, sehingga *cluster* ini terdiri dari (16, 17, 20, 33).
21. Langkah 21, *cluster* ke-6 digabungkan dengan *cluster* ke-10, sehingga *cluster* ini terdiri dari (25, 30, 31, 35).
22. Langkah 22, objek ke-24 digabungkan dengan *cluster* ke-3, sehingga *cluster* ini terdiri dari (1, 3, 19, 21, 24).
23. Langkah 23, objek ke-26 digabungkan dengan *cluster* ke-12, sehingga *cluster* ini terdiri dari (13, 26, 27).

24. Langkah 24, *cluster* ke-10 digabungkan dengan *cluster* ke-14, sehingga *cluster* ini terdiri dari (32, 34, 37, 38).
25. Langkah 25, *cluster* ke-1 digabungkan dengan *cluster* ke-3, sehingga *cluster* ini terdiri dari (1, 2, 3, 10, 19, 21, 22, 24, 28).
26. Langkah 26, objek ke-36 digabungkan dengan *cluster* ke-10, sehingga *cluster* ini terdiri dari (25, 30, 31, 35, 36).
27. Langkah 27, objek ke-11 digabungkan dengan *cluster* ke-13, sehingga *cluster* ini terdiri dari (11, 12, 23).
28. Langkah 28, *cluster* ke-4 digabungkan dengan *cluster* ke-3, sehingga *cluster* ini terdiri dari (1, 2, 3, 4, 6, 10, 18, 19, 21, 22, 24, 28).
29. Langkah 29, objek ke-8 digabungkan dengan *cluster* ke-13, sehingga *cluster* ini terdiri dari (9, 11, 12, 23, 29).
30. Langkah 30, *cluster* ke-11 digabung dengan *cluster* ke-15, sehingga *cluster* ini terdiri dari (5, 7, 8, 14).
31. Langkah 31, objek ke-15 digabungkan dengan *cluster* ke-10, sehingga *cluster* ini terdiri dari (15, 25, 30, 31, 35, 36).
32. Langkah 32, *cluster* ke-7 digabungkan dengan *cluster* ke-14, sehingga *cluster* ini terdiri dari (16, 17, 20, 32, 33, 34, 37, 38).
33. Langkah 33, *cluster* ke-10 digabungkan dengan *cluster* ke-14, sehingga *cluster* ini terdiri dari (15, 16, 17, 20, 25, 30, 31, 32, 33, 34, 35, 36, 37, 38).
34. Langkah 34, *cluster* ke-13 digabung dengan *cluster* ke-15, sehingga *cluster* ini terdiri dari (5, 7, 8, 9, 11, 12, 14, 23, 29).
35. Langkah 35, *cluster* ke-12 digabungkan dengan *cluster* ke-15, sehingga *cluster* ini terdiri dari (5, 7, 8, 9, 11, 12, 13, 14, 23, 26, 27, 29).

36. Langkah 36, *cluster* ke-3 digabungkan dengan *cluster* ke-14, sehingga *cluster* ini terdiri dari (1, 2, 3, 4, 6, 10, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38).
37. Langkah 37, *cluster* ke-14 digabungkan dengan *cluster* ke-15, sehingga *cluster* ini terdiri dari (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38).



Gambar 4.5 *Dendrogram Complete Linkage*

Representasi visual dari langkah-langkah penggabungan digambarkan dalam bentuk *dendrogram* pada Gambar 4.5.

Setelah perhitungan berakhir, berikutnya menentukan validitas setiap *cluster* menggunakan *multiscale bootstrap*. Metode *multiscale bootstrap* adalah metode yang digunakan untuk mengukur validitas dalam *clustering* hierarki, dengan menggunakan pendekatan *multiscale bootstrap* dapat memperoleh estimasi ketidakpastian yang lebih akurat dalam hasil *clustering* hierarki. Struktur dalam *clustering* hierarki dibangun secara bertahap dengan menggabungkan pasangan

cluster berdasarkan jarak antar mereka. Namun, proses ini dapat menghasilkan variasi dalam hasil *clustering* tergantung pada data yang digunakan. Oleh karena itu, metode *multiscale bootstrap* mengatasi ketidakpastian dengan melakukan repetisi *bootstrap* pada subset data yang diambil pada skala yang berbeda. Dengan *me-resampling* data pada skala yang beragam, metode ini dapat menghasilkan sejumlah replika *clustering* hierarki yang mencerminkan variasi dalam hasil *clustering*.

Perhitungan *multiscale bootstrap* pada masing-masing *cluster* yang pertama adalah memisalkan sampel data pengamatan (N), Berdasarkan data penelitian yang penulis gunakan, jumlah data pengamatan yaitu Kabupaten/Kota di Jawa Timur sebanyak $N = 38$, dan nilai dari jumlah data pengamatan *bootstrap* untuk $N' = 19, 25, 31, 38, 44, 50$. Menggunakan 6 skala dapat memberikan representasi yang cukup luas dari berbagai tingkat resolusi atau skala yang ada dalam data. Rentang resolusi yang dicakup oleh skala-skala tersebut dapat memberikan informasi yang komprehensif tentang pola *cluster* dalam data. Analisis *multiscale bootstrap* dilakukan dengan menggunakan paket *pvclust* dengan bantuan *software* R pada data indikator pendidikan SMA sederajat di Jawa Timur tahun 2021.

Pada pendekatan *multiscale bootstrap* terdapat nilai *approximately unbiased* (AU) dan nilai *bootstrap probability* (BP). Nilai AU dan BP digunakan untuk menunjukkan signifikansi statistik. Nilai AU memiliki keunggulan ketidakbiasannya yang lebih mendekati nilai sebenarnya dikarenakan menggunakan *resampling* pada beberapa skala yang berbeda (*multiscale*), sedangkan nilai BP menggunakan *resampling* pada satu skala. Untuk menentukan nilai signifikansi AU setiap *cluster* diperlukan nilai koefisien v dan c dari hasil

perhitungan $z(N')$. Nilai $z(N')$ adalah nilai $z = -\Phi^{-1}(BP)$, perhitungan nilai v dan c dengan menyesuaikan kurva teoritis, menggunakan rumus perhitungan (2.5).

Berikut ini adalah langkah-langkah perhitungan nilai AU menggunakan metode *multiscale bootstrap* berdasarkan metode *complete linkage*. Gambar 4.6 menunjukkan plot diagnostik *multiscale bootstrap* untuk langkah penggabungan 1. Terdapat nilai z yang diamati, diplot untuk $\sqrt{N'/N}$, nilai AU, nilai BP, nilai v dan c . Contoh pertama untuk menghitung nilai AU pada langkah 1 yang anggotanya terdiri dari Kabupaten Banyuwangi dan Kabupaten Bojonegoro dengan menyesuaikan kurva pada Gambar 4.6 untuk mengestimasi nilai v dan c .

$$\begin{aligned}
 z(N') &= v \sqrt{\frac{N'}{N}} + c \sqrt{\frac{N}{N'}} \\
 0,737505 &= v \sqrt{\frac{19}{38}} + c \sqrt{\frac{38}{19}} \\
 0,445746 &= v \sqrt{\frac{25}{38}} + c \sqrt{\frac{38}{25}} \\
 \hline
 0,737505 &= v \, 0,70711 + c \, 1,41421 & \left| \begin{array}{l} * 1,23288 \\ * 1,41421 \end{array} \right. \\
 0,445746 &= v \, 0,81111 + c \, 1,23288 \\
 \hline
 0,909256 &= v \, 0,87178 + c \, 1,74355 \\
 0,630379 &= v \, 1,14708 + c \, 1,74355 \\
 \hline
 0,278877 &= -0,275298 v \\
 v &= \frac{0,278877}{-0,275298} \\
 v &= -1,013
 \end{aligned}$$

Selanjutnya, menghitung nilai c dengan mensubstitusi nilai v :

$$\begin{aligned}
 0,737505 &= v \, 0,70711 + c \, 1,41421 \\
 0,737505 &= (-1,013) \, 0,70711 + c \, 1,41421 \\
 0,737505 &= (-0,716302) + c \, 1,41421
 \end{aligned}$$

$$1,453807 = c \cdot 1,41421$$

$$c = 1,028$$

Berdasarkan perhitungan di atas diperoleh nilai $v = -1,013$ dan $c = 1,028$.

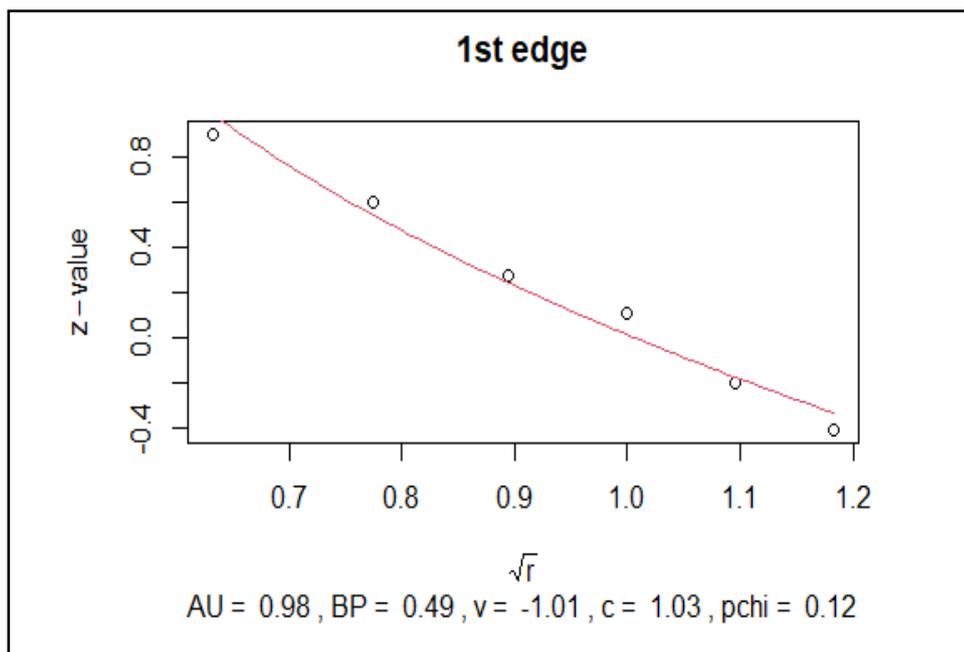
Kemudian menghitung nilai AU untuk *cluster* 1 menggunakan rumus (2.6):

$$AU = \Phi(-(-1,013) + 1,028)$$

$$AU = \Phi(2,041)$$

$$AU = 0,97937 \sim 0,98$$

Sehingga, nilai AU untuk *cluster* 1 sebesar 0,98 yang berarti lebih besar sama dengan 0,95 maka dapat disimpulkan bahwa suatu *cluster* valid. Hasil perhitungan lengkap pada semua *cluster* dapat dilihat pada Tabel 4.4.



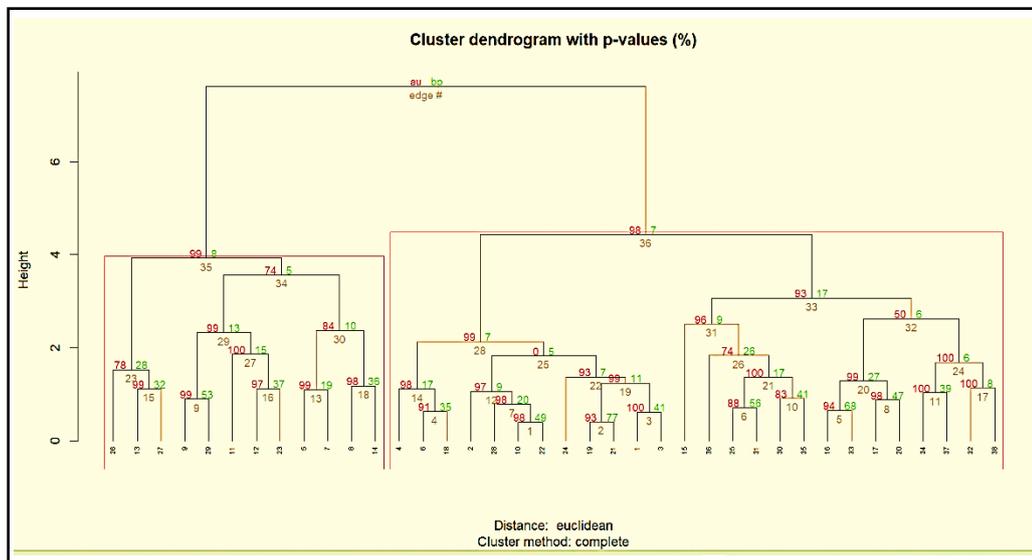
Gambar 4.6 Plot Diagnostik dari *Multiscale Bootstrap* untuk Langkah Penggabungan 1

Tabel 4.4 Hasil Perhitungan AU, BP, v dan c dengan Metode *Multiscale Bootstrap*

Edge	AU	BP	v	c
1.	0,979	0,494	-1.013	1,028
2.	0,934	0,770	-1.122	0,383
3.	1.000	0,406	-1.935	2,174
4.	0,906	0,345	-0,46	0,859
5.	0,937	0,679	-0,997	0,532
6.	0,876	0,557	-0,65	0,506
7.	0,980	0,199	-0,601	1,446
8.	0,985	0,471	-1.047	1,120
9.	0,989	0,529	-1.179	1,106
10.	0,835	0,410	-0,374	0,600
11.	0,999	0,393	-1.488	1,759
12.	0,971	0,087	-0,271	1,633
13.	0,985	0,190	-0,653	1,529
14.	0,976	0,169	-0,509	1,466
15.	0,988	0,322	-0,892	1,354
16.	0,972	0,371	-0,789	1,117
17.	1.000	0,083	-1.280	2,661
18.	0,982	0,360	-0,874	1,234
19.	0,990	0,108	-0,545	1,782
20.	0,994	0,270	-0,937	1,549
21.	0,998	0,174	-0,999	1,937
22.	0,934	0,072	-0,021	1,482
23.	0,784	0,284	-0,109	0,678
24.	0,998	0,060	-0,700	2,252
25.	0,004	0,045	2.175	-0,48
26.	0,735	0,264	0,001	0,629
27.	0,998	0,146	-0,934	1,989
28.	0,989	0,066	-0,383	1,892
29.	0,986	0,126	-0,530	1,677
30.	0,845	0,103	0,127	1,140
31.	0,960	0,086	-0,193	1,562
32.	0,499	0,061	0,774	0,772
33.	0,927	0,174	-0,258	1,196
34.	0,743	0,054	0,476	1,128
35.	0,988	0,076	-0,418	1,854
36.	0,981	0,074	-0,309	1,759
37.	1.000	1.000	0	0

Tabel 4.5 Nilai Koefisien dan Selisih Jarak *Complete Linkage*

Edge	Koefisien	Selisih
1	0,40	0,01
2	0,41	0,21
3	0,62	0,01
4	0,63	0,03
5	0,66	0,05
6	0,71	0,08
7	0,79	0,08
8	0,87	0,03
9	0,90	0,01
10	0,91	0,12
11	1,03	0,03
12	1,06	0,04
13	1,10	0,02
14	1,12	0,00
15	1,12	0,00
16	1,12	0,02
17	1,14	0,04
18	1,18	0,06
19	1,24	0,05
20	1,29	0,08
21	1,37	0,00
22	1,37	0,16
23	1,53	0,16
24	1,69	0,13
25	1,82	0,02
26	1,84	0,02
27	1,86	0,27
28	2,13	0,20
29	2,33	0,04
30	2,37	0,14
31	2,51	0,11
32	2,62	0,45
33	3,07	0,50
34	3,57	0,10
35	3,67	0,76
36	4,43	3,19
37	7,62	



Gambar 4.8 Hasil *Dendrogram* Metode *Hierarchical Clustering Multiscale Bootstrap*

4.8 Karakterisasi Hasil *Cluster*

Setelah diketahui banyaknya *cluster* yang diperoleh, selanjutnya akan diberikan karakter masing-masing *cluster* serta nama Kabupaten/Kota ke dalam *cluster*, yang diringkas dalam bentuk Tabel 4.6. Pada Tabel 4.6 diketahui bahwa pengelompokan hierarki Kabupaten/Kota di Jawa Timur dengan metode *hierarchical clustering multiscale bootstrap* terdiri dari 2 *cluster*. *Cluster* pertama terdiri dari 26 Kabupaten/Kota, *cluster* kedua terdiri dari 12 Kabupaten/Kota.

Tabel 4.6 Jumlah *Cluster* dan Anggota dari Setiap *Cluster* dengan Metode *Hierarchical Clustering Multiscale Bootstrap*

<i>Cluster</i>	Jumlah Anggota	Kabupaten/Kota	Nilai AU	Keterangan
1	26	Kabupaten Tulungagung, Kabupaten Kediri, Kabupaten Nganjuk, Kabupaten Ponorogo, Kabupaten Pamekasan, Kabupaten Banyuwangi, Kabupaten Bojonegoro, Kabupaten Lamongan, Kabupaten Madiun, Kabupaten Ngawi, Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Sidoarjo, Kota Madiun, Kabupaten Gresik, Kota Blitar, Kota Kediri, Kota Mojokerto, Kabupaten Mojokerto, Kota Probolinggo, Kabupaten Jombang, Kabupaten Magetan, Kota Pasuruan, Kota Surabaya, Kota Malang, Kota Batu.	0,98	Valid
2	12	Kabupaten Bangkalan, Kabupaten Probolinggo, Kabupaten Sampang, Kabupaten Jember, Kabupaten Sumenep, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Tuban, Kabupaten Blitar, Kabupaten Malang, Kabupaten Lumajang, Kabupaten Pasuruan.	0,99	Valid

Salah satu cara yang dapat digunakan untuk mengetahui karakteristik setiap *cluster* yaitu dengan menghitung *mean* dari masing-masing variabel berdasarkan *cluster*. Hasil perhitungan *mean* karakteristik *cluster* dapat dilihat dalam Tabel 4.7.

Tabel 4.7 Hasil Perhitungan Nilai *Mean* pada Masing-Masing Variabel berdasarkan *Cluster*

<i>Cluster</i>	X_1	X_2	X_3	X_4	X_5
1	93,03	81,70	71,3	28,62	95,34
2	75,3	64,42	51,8	17,61	88,92

Berdasarkan Tabel 4.7 diketahui bahwa *cluster* 1 merupakan daerah-daerah yang memiliki penduduk dengan karakteristik rata-rata APK, APS, APM, pendidikan SMA sederajat yang ditamatkan dan AMH yang lebih tinggi daripada *cluster* 2 dengan selisih yang cukup tinggi. Selisih rata-rata pada variabel APK sebanyak 17,73%, APS 17,28%, APM 19,5%, pendidikan SMA sederajat yang ditamatkan 11,01% dan AMH 6,42%. *Cluster* 1 terdiri dari 26 daerah diantaranya Kabupaten Tulungagung, Kabupaten Kediri, Kabupaten Nganjuk, Kabupaten Ponorogo, Kabupaten Pamekasan, Kabupaten Banyuwangi, Kabupaten Bojonegoro, Kabupaten Lamongan, Kabupaten Madiun, Kabupaten Ngawi, Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Sidoarjo, Kota Madiun, Kabupaten Gresik, Kota Blitar, Kota Kediri, Kota Mojokerto, Kabupaten Mojokerto, Kota Probolinggo, Kabupaten Jombang, Kabupaten Magetan, Kota Pasuruan, Kota Surabaya, Kota Malang dan Kota Batu.

Sedangkan pada *cluster* 2 beranggotakan 12 daerah yaitu Kabupaten Bangkalan, Kabupaten Probolinggo, Kabupaten Sampang, Kabupaten Jember, Kabupaten Sumenep, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Tuban, Kabupaten Blitar, Kabupaten Malang, Kabupaten Lumajang dan Kabupaten

Pasuruan. Nilai AU pada *cluster* 1 sebesar 0,98 dan 0,99 menandakan bahwa kedua *cluster* memiliki nilai $AU \geq 0,95$ yang artinya kedua *cluster* dapat dikatakan valid.

Cluster 1 memiliki ukuran tingkat partisipasi SMA sederajat lebih banyak daripada *cluster* 2 dikarenakan perbedaan infrastruktur perkotaan dan pedesaan yang diduga menjadi salah satu penyebab terjadinya perbedaan penduduk dalam pencapaian tingkat pendidikan. Keberadaan fasilitas dan kemudahan mengakses sarana dan prasarana pendidikan di perkotaan menjadi daya tarik penduduk untuk mengenyam pendidikan. Hal ini menyebabkan tingkat pendidikan penduduk di daerah perkotaan dan sekitarnya lebih tinggi dibandingkan dengan daerah lainnya (BPS, 2021).

Misalnya pada daerah *cluster* 1 yaitu Kota metropolitan Surabaya yang berbatasan langsung dengan Kabupaten Sidoarjo, Kota Mojokerto, dan Kabupaten Gresik memiliki persentase variabel angka partisipasi lebih banyak dibandingkan dengan daerah-daerah pada *cluster* 2 seperti Kabupaten Bangkalan, Kabupaten Probolinggo, Kabupaten Sampang dikarenakan penduduk di daerah perkotaan lebih mudah mengakses pendidikan yang lebih bervariasi. Selain itu, sektor perekonomian di Kota Surabaya yang dominan pada sektor tersier, umumnya mensyaratkan minimal pendidikan pada jenjang menengah atas (BPS, 2021). Contoh lain Kabupaten Bangkalan yang memiliki APS terendah di Jawa Timur menandakan bahwa tingkat partisipasi penduduk pada daerah tersebut perlu ditingkatkan. Oleh karena itu untuk mencapai tujuan pemerintah daerah, khususnya pemerintah provinsi Jawa Timur dalam menyelenggarakan pendidikan yang lebih mudah diakses oleh penduduk, seperti menyediakan sarana dan prasarana yang memadai agar mendorong tingkat partisipasi sekolah (BPS, 2021).

Pendidikan merupakan bagian dari proses perjalanan yang dilakukan manusia agar menjadi manusia yang berkualitas. Agar mencapai tujuan tersebut maka manusia diharuskan menuntut ilmu dan memiliki pengetahuan untuk bekal kehidupannya. Salah satu cara agar manusia mendapatkan ilmu pengetahuan adalah dengan membaca, seperti yang telah penulis paparkan dalam subbab 2.9 dalam surah Al-‘Alaq ayat 1-5, Allah memerintahkan untuk membaca pada ayat 1 dan 3. Perintah membaca pertama dimaksudkan untuk belajar sesuatu yang belum diketahui, sedangkan yang kedua perintah untuk mengajarkan ilmu kepada orang lain. Hal ini mengindikasikan bahwa dalam proses belajar dan pembelajaran dituntut adanya usaha yang maksimal dengan memfungsikan segala komponen berupa alat-alat potensial yang ada pada diri manusia. Setelah mendapatkan ilmu melalui pendidikan, maka amanat selanjutnya adalah mengajarkan ilmu pengetahuan tersebut dengan cara tetap memfungsikan segala potensi yang ada. Dijelaskan juga tentang penciptaan manusia dari segumpal darah yang dapat menciptakan manusia dengan kelebihan dan kekurangan masing-masing. Manusia merupakan makhluk ciptaan Allah dengan kelebihan diberikannya akal sehingga manusia dapat berfikir dengan adanya akal tersebut. Manusia diperintah untuk selalu mengingat Allah ketika mencari ilmu, dengan tujuan agar dipermudah dalam mencari ilmu dan ilmu tersebut dapat menjadi sebuah keberkahan bagi dirinya.

Salah satu faktor adanya negara yang berkualitas ialah dikarenakan di dalamnya terdapat manusia yang berkualitas pula. Oleh karena itu, mewujudkan manusia yang berkualitas dengan berbagai metode yang efektif perlu dilakukan salah satunya melalui peningkatan kualitas pendidikan. Menjadikan manusia yang berpendidikan juga telah diupayakan oleh pemerintah, salah satunya dengan

menerapkan pendidikan wajib belajar. Pendidikan wajib belajar 12 tahun yaitu program yang dicanangkan oleh pemerintah untuk warga negara agar menempuh pendidikan hingga SMA sederajat. Kondisi pendidikan SMA sederajat di Jawa Timur masih membutuhkan pemerataan dikarenakan pada beberapa Kabupaten masih memiliki ketimpangan indikator pendidikan tingkat SMA sederajat. Seperti pada Kabupaten Bangkalan, Kabupaten Probolinggo, Kabupaten Sampang, Kabupaten Jember, Kabupaten Sumenep, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Tuban, Kabupaten Blitar, Kabupaten Malang, Kabupaten Lumajang dan Kabupaten Pasuruan yang memiliki tingkat partisipasi kasar, sekolah, murni, pendidikan SMA sederajat yang ditamatkan, angka melek huruf yang rendah. Salah satu upaya yang dapat diterapkan demi mewujudkan pemerataan kondisi pendidikan tersebut di suatu daerah adalah mengelompokkan daerah-daerah yang memiliki karakteristik yang sama, sehingga proses peningkatan dan pemerataan dapat sesuai dengan kondisi daerah masing-masing. Beberapa manfaat dari *clustering* tersebut adalah untuk memprioritaskan perbaikan dan memberikan gambaran kondisi pendidikan, sehingga mempermudah dan mempercepat proses penanganan.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dijabarkan, maka dapat disimpulkan bahwa hasil dari penerapan metode *hierarchical clustering multiscale bootstrap* pada indikator pendidikan tingkat SMA sederajat di Provinsi Jawa Timur tahun 2021 yaitu memperoleh 2 *cluster* valid dengan nilai $AU \geq 0,95$. *Cluster* 1 terdiri dari 26 Kabupaten/Kota yaitu Kabupaten Tulungagung, Kabupaten Kediri, Kabupaten Nganjuk, Kabupaten Ponorogo, Kabupaten Pamekasan, Kabupaten Banyuwangi, Kabupaten Bojonegoro, Kabupaten Lamongan, Kabupaten Madiun, Kabupaten Ngawi, Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Sidoarjo, Kota Madiun, Kabupaten Gresik, Kota Blitar, Kota Kediri, Kota Mojokerto, Kabupaten Mojokerto, Kota Probolinggo, Kabupaten Jombang, Kabupaten Magetan, Kota Pasuruan, Kota Surabaya, Kota Malang dan Kota Batu dengan nilai AU 0,98, *cluster* 1 dikategorikan sebagai daerah-daerah dengan APK, APS, APM, Pendidikan SMA sederajat yang ditamatkan dan AMH yang lebih tinggi daripada *cluster* 2 dengan selisih rata-rata setiap variabelnya yang cukup tinggi. Sedangkan pada *cluster* 2 terdiri dari 12 Kabupaten di antaranya Kabupaten Bangkalan, Kabupaten Probolinggo, Kabupaten Sampang, Kabupaten Jember, Kabupaten Sumenep, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Tuban, Kabupaten Blitar, Kabupaten Malang, Kabupaten Lumajang dan Kabupaten Pasuruan dengan nilai AU 0,99. Kedua *cluster* tersebut menunjukkan bahwa *cluster* mengandung keakuratan sebesar 98% dan 99%.

5.2 Saran

Pada penelitian ini peneliti hanya membahas metode *hierarchical clustering multiscale bootstrap* dan diterapkan pada data indikator pendidikan tingkat SMA sederajat di Jawa Timur. Oleh karena itu peneliti berharap untuk penelitian selanjutnya dapat menerapkan analisis hierarki yang lain maupun metode non hierarki serta diterapkan pada data/objek yang lain.

DAFTAR PUSTAKA

- Al-Mahalli Jamaluddin & Al-Suyuthi Jamaluddin. (2007). *Tafsir Jalalain: Tafsir Alqur'an Al-'Adzim*. Surabaya: Al-Haramain Jaya Indonesia.
- Anuraga, G. (2015). Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokan Kemiskinan di Jawa Timur. *Jurnal Statistika*, 03(1), 27–33. <https://jurnal.unimus.ac.id/index.php/statistik/article/download/1432/1485>
- Badan Perencanaan Pembangunan Nasional. (2015). *Seri Analisis Pembangunan Wilayah Provinsi Jawa Timur 2015*.
- BPS. (2020). Statistik Pendidikan Provinsi Jawa Timur 2020. *Badan Pusat Statistik Provinsi Jawa Timur*. <https://jatim.bps.go.id/publication/2021/12/06/143ff261cca315e5cbfb82b5/statistik-pendidikan-provinsi-jawa-timur-2020.html>
- BPS. (2021). Statistik Pendidikan Provinsi Jawa Timur 2021. *Badan Pusat Statistik Provinsi Jawa Timur*. <https://jatim.bps.go.id/publication/2022/11/30/e4718c89255814d14cd41703/statistik-pendidikan-provinsi-jawa-timur-2021.html>
- BPS. (2021). *Angka Partisipasi Murni (APM) dan Angka Partisipasi Kasar (APK) Menurut Jenjang Pendidikan di Jawa Timur. Survei Sosial Ekonomi Nasional (SUSENAS) Tahun 2021*.
- Dabhi, D. P., & Patel, M. R. (2016). Extensive Survey on Hierarchical Clustering Methods in Data Mining. *International Research Journal of Engineering and Technology (IRJET)*, 03(11), 659–665. <https://www.irjet.net/archives/V3/i11/IRJET-V3I11115>
- Departemen Agama RI, (2006). *Al-Qur'an Al-Karim dan Terjemah Bahasa Indonesia*. Kudus: Menara Kudus.
- Dhira, O. L., & Atika, F. (2019). Analisis Klaster untuk Mengidentifikasi Kualitas Pendidikan Sekolah Menengah Sederajat di Indonesia. *Jurnal Ilmiah Populer*, 1, 12-20. <http://median.bpsjambi.id/index.php/median/article/view/18>
- Effron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. United States of America: Chapman and Hall.
- Ekananda, M. (2015). *Ekonometrika Dasar untuk Penelitian Dibidang Ekonomi, Sosial dan Bisnis*. Mitra Wacana Media.

- Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). In *Atmospheric Pollution Research* (Vol. 11, Issue 1). Turkish National Committee for Air Pollution Research and Control.
<https://doi.org/10.1016/j.apr.2019.09.009>
- Gudono. (2011). *Analisis Data Multivariat*. BPFE-YOGYAKARTA.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometric (5th Ed)*. New York: McGraw-Hill.
- Hair, J. F. C., Black, W. J., Babin, B., & Anderson, R. E. (2010). *Multivariate Data Analysis (7th Edition)*. United States: Pearson Prentice Hall.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Techniques (3rd edition)*. Waltham: Morgan Kaufmann Publishers.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms For Clustering Data*. New Jersey: Pearson Prentice Hall.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.
- Kapita, S. N., & Abdullah, S. Do. (2020). Pengelompokan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-SOM. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 56–61. <https://doi.org/10.33387/jiko>
- Kaiser, H. (1974). An index of factor simplicity. *Psychometrika*, 31-36.
- Kaufman, L., & Rousseeuw, P. (1990). *Finding Groups In Data An Introduction to Cluster Analysis*. Canada: John Wiley & Sons.
- Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *Multivariate Calibration and Classification*. NIR Publications.
- Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 13 Tahun 2022 tentang *Perubahan Atas Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 22 Tahun 2020 Tentang Rencana Strategis Kementerian Pendidikan Dan Kebudayaan Tahun 2020-2024*.
- Peraturan Pemerintah Republik Indonesia Nomor 47 Tahun 2008 tentang *Wajib Belajar* (Lembaran Negara Republik Indonesia Tahun 2008 Nomor 90, Tambahan Lembaran Negara Republik Indonesia Nomor 4863).
- Pribadi, W. W., Yunus, A., & Wiguna, S. A. (2022). Perbandingan Metode K-Means

- Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang. *Jurnal Mahasiswa Teknik Informatika*, 6(2), 493–500. <https://ejournal.itn.ac.id/index.php/jati/article/download/4808/3352/>
- Ramly, M. M., & Poerwanto, B. (2022). Analisis Hierarchical Clustering Multiscale Bootstrap (Kasus : Indikator Kemiskinan di Provinsi Sulawesi Selatan Tahun 2020). *Journal of statistics and its application on teaching and research*, 4(3), 142–152. <https://doi.org/10.35580/variansiunm26>
- Ros, F., & Guillaume, S. (2019). A hierarchical clustering algorithm and an improvement of the single linkage criterion to deal with noise. *Expert Systems with Applications*, 128, 96–108. <https://doi.org/10.1016/j.eswa.2019.03.031>
- Safarina, Y., Hayati, M. N., & Nasution, Y. N. (2019). Penerapan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Indeks Pembangunan Manusia (IPM) di Kalimantan Timur. *Prosiding Seminar Nasional Matematika, Statistika Dan Aplikasinya*, (1), 29–35. <http://jurnal.fmipa.unmul.ac.id/index.php/SNMSA/article/view/523>
- Sharuddin, S., & Muda, N. (2015). Performance Comparison between Bootstrap and Multiscale Bootstrap for Assessing Phylogenetic Tree for RNA polymerase. *Sains Malaysiana*, 44(11), 1643-1651. <http://journalarticle.ukm.my/9374/>
- Qutub, S. (2011). Sumber-Sumber Ilmu Pengetahuan dalam Al-Qur'an dan Hadits. *Jurnal Humaniora*, 2(2), 1339-1350. <https://doi.org/10.21512/humaniora.v2i2.3198>
- Shihab, M. Q. (2009). *Tafsir Al-Mishbah (Pesan, Kesan dan Kekeragaman Al-Qur'an)*. Jakarta: Lentera Hati.
- Supranto, J. (2004). *Analisis Multivariat: Arti & Interpretasi*. Jakarta: PT Rineka Cipta.
- Suzuki, R., & Shimodaira, H. (2004). An Application of Multiscale Bootstrap Resampling to Hierarchical Clustering of Microarray Data: How Accurate are These Clusters?. *In proceedings by the Fifteenth International Conference on Genome Informatics (GIW 2004)*, p. P034.
- Laksmi, A. D., Kusnandar, D., & Imro'ah, N. (2023). Hierarchical Clustering Multiscale Bootstrap pada Capaian Dimensi IPM Tahun 2020 yang

- Terdampak Covid-19. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 12(02), 185-194
<https://jurnal.untan.ac.id/index.php/jbmstr/article/view/65277>
- Shimodaira, H. (2004). Approximately Unbiased Tests of Regions Using Multistep-Multiscale Bootstrap Resampling. *The Annals of Statistics*, 32(06), 2616-2641
<https://doi.org/10.1214/009053604000000823>
- Suzuki, R., & Shimodaira, H. (2006). Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540–1542.
<https://doi.org/10.1093/bioinformatics/btl117>
- Suzuki, R., & Shimodaira, H. (2006). Hierarchical Clustering with P-values via Multiscale Bootstrap Resampling.
<https://www.researchgate.net/publication/230710851>
- Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining - Instructor's Solution Manual*. Boston: Pearson Education.
- Undang-Undang Dasar Negara Republik Indonesia Tahun 1945*. Pasal 28C ayat 1.
<https://www.bkn.go.id/wp-content/uploads/2014/06/UUD-1945.pdf>
- Whendasmoro R. G., & Joseph, J. (2022). Analisis Penerapan Normalisasi Data dengan Menggunakan Z-score pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872-876.
<https://doi.org/10.30865/jurikom.v9i4.4526>
- Wijaya, T., & Budiman, S. (2016). *Analisis Multivariat Untuk Penelitian Manajemen*. Yogyakarta: Pohon Cahaya.

LAMPIRAN

Lampiran 1. Data Indikator Pendidikan SMA Sederajat di Jawa Timur Tahun 2021

No.	Kabupaten/Kota	Angka Partisipasi Kasar	Angka Partisipasi Sekolah	Angka Partisipasi Murni	Pendidikan SMA yang ditamatkan	Rata-Rata Lama Sekolah	Angka Melek Huruf
1	Kabupaten Pacitan	88,36	76,94	66,97	17,27	7,82	92,66
2	Kabupaten Ponorogo	81,75	80,74	68,84	21,83	7,93	88,71
3	Kabupaten Trenggalek	90,30	77,96	69,72	19,95	7,97	94,69
4	Kabupaten Tulungagung	69,64	79,99	62,82	23,95	8,65	96,36
5	Kabupaten Blitar	81,45	70,46	60,69	20,17	7,96	94,23
6	Kabupaten Kediri	82,57	81,84	68,77	25,81	8,50	94,68
7	Kabupaten Malang	93,32	65,48	55,52	17,94	7,84	95,29
8	Kabupaten Lumajang	63,68	55,42	46,50	16,99	7,65	94,50
9	Kabupaten Jember	96,13	69,83	60,92	17,37	6,78	87,38
10	Kabupaten Banyuwangi	80,68	79,73	61,97	22,55	7,94	91,55
11	Kabupaten Bondowoso	70,38	71,24	49,87	16,88	6,57	81,69
12	Kabupaten Situbondo	79,90	65,56	58,52	20,36	7,19	84,91
13	Kabupaten Probolinggo	66,62	60,37	38,45	17,34	6,76	86,63
14	Kabupaten Pasuruan	72,37	64,68	45,54	20,68	7,61	94,05
15	Kabupaten Sidoarjo	125,9	87,45	70,98	36,87	10,83	99,12
16	Kabupaten Mojokerto	88,16	82,34	75,60	30,48	9,06	95,64
17	Kabupaten Jombang	95,94	90,39	73,70	26,20	8,84	96,26
18	Kabupaten Nganjuk	76,68	83,19	65,35	24,53	8,22	93,10
19	Kabupaten Madiun	92,90	78,70	74,08	25,00	8,03	91,40
20	Kabupaten Magetan	100,93	88,36	80,33	26,71	8,69	93,72
21	Kabupaten Ngawi	92,27	78,45	73,62	21,87	7,97	91,29
22	Kabupaten Bojonegoro	78,27	80,22	63,57	21,94	7,93	93,10
23	Kabupaten Tuban	70,93	65,55	56,31	21,30	7,94	89,21
24	Kabupaten Lamongan	92,18	71,77	67,63	26,28	8,50	94,92
25	Kabupaten Gresik	107,29	85,67	80,48	33,55	9,86	97,44
26	Kabupaten Bangkalan	57,00	51,28	41,61	13,80	6,66	89,58
27	Kabupaten Sampang	62,54	57,29	41,78	11,90	5,91	83,26
28	Kabupaten Pamekasan	80,84	75,91	62,52	20,26	7,58	90,25
29	Kabupaten Sumenep	89,32	75,96	65,94	16,67	6,84	86,32
30	Kota Kediri	95,57	87,04	80,60	37,87	11,21	97,98
31	Kota Blitar	106,09	91,17	85,02	34,26	10,98	98,62
32	Kota Malang	101,01	84,30	65,22	33,39	11,05	98,49
33	Kota Probolinggo	96,99	82,67	72,73	31,25	9,75	95,59
34	Kota Pasuruan	87,50	75,59	63,45	34,49	10,20	98,21
35	Kota Mojokerto	103,69	80,02	78,92	39,26	11,10	98,47
36	Kota Madiun	95,97	96,82	80,92	43,05	11,93	99,07
37	Kota Surabaya	99,52	69,97	66,57	34,77	10,52	99,03
38	Kota Batu	107,83	76,98	73,31	30,84	9,90	98,62

Lampiran 2. Hasil Standardisasi Data

No.	Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆
1	0,06	0,07	0,16	-1,01	-0,51	-0,14
2	-0,39	0,44	0,31	-0,43	-0,44	-0,98
3	0,19	0,17	0,39	-0,67	-0,41	0,29
4	-1,21	0,36	-0,20	-0,15	0,03	0,65
5	-0,41	-0,56	-0,38	-0,64	-0,42	0,19
6	-0,33	0,54	0,31	0,08	-0,06	0,29
7	0,40	-1,04	-0,82	-0,93	-0,50	0,42
8	-1,61	-2,02	-1,58	-1,05	-0,62	0,25
9	0,59	-0,62	-0,36	-1,00	-1,19	-1,26
10	-0,46	0,34	-0,27	-0,33	-0,43	-0,37
11	-1,16	-0,48	-1,29	-1,06	-1,33	-2,47
12	-0,51	-1,03	-0,56	-0,61	-0,92	-1,78
13	-1,42	-1,54	-2,26	-1,00	-1,20	-1,42
14	-1,02	-1,12	-1,66	-0,57	-0,65	0,16
15	2,62	1,09	0,50	1,51	1,46	1,23
16	0,05	0,59	0,89	0,68	0,30	0,49
17	0,58	1,37	0,73	0,14	0,16	0,62
18	-0,73	0,67	0,02	-0,08	-0,25	-0,05
19	0,37	0,24	0,76	-0,02	-0,37	-0,41
20	0,92	1,17	1,29	0,20	0,06	0,09
21	0,33	0,21	0,72	-0,42	-0,41	-0,43
22	-0,62	0,38	-0,13	-0,41	-0,44	-0,05
23	-1,12	-1,04	-0,75	-0,49	-0,43	-0,87
24	0,32	-0,43	0,21	0,15	-0,06	0,34
25	1,35	0,91	1,30	1,08	0,83	0,87
26	-2,07	-2,42	-1,99	-1,46	-1,27	-0,79
27	-1,69	-1,84	-1,98	-1,70	-1,76	-2,13
28	-0,45	-0,03	-0,22	-0,63	-0,67	-0,65
29	0,13	-0,03	0,07	-1,09	-1,15	-1,48
30	0,55	1,05	1,31	1,63	1,71	0,99
31	1,27	1,45	1,69	1,17	1,56	1,13
32	0,92	0,78	0,01	1,06	1,61	1,10
33	0,65	0,62	0,64	0,78	0,75	0,48
34	0,00	-0,06	-0,14	1,20	1,05	1,04
35	1,11	0,37	1,17	1,81	1,64	1,09
36	0,58	1,99	1,34	2,30	2,18	1,22
37	0,82	-0,61	0,12	1,24	1,26	1,21
38	1,39	0,07	0,69	0,73	0,85	1,13

Lampiran 3. Tabel Jarak *Euclidean*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0,00	1,19	0,62	1,78	1,07	1,34	1,61	3,21	1,51	1,02	3,05	2,22	3,50	2,48	3,98	2,02	2,05	1,38	1,24	2,18	0,92	1,01	2,06	1,37	3,01	4,00	3,96	0,91	1,35	3,29	3,47	2,66	2,12	2,53	3,42	4,26	2,81	2,58
2	1,19	0,00	1,44	1,91	1,7	1,37	2,51	3,61	1,71	0,85	2,58	1,90	3,48	2,83	4,25	1,98	2,20	1,12	1,14	2,17	1,01	1,06	1,96	1,82	3,15	4,17	3,88	0,81	1,11	3,22	3,55	2,90	2,19	2,70	3,5	4,08	3,18	3,03
3	0,62	1,44	0,00	1,66	1,22	1,00	1,74	3,47	1,96	1,20	3,58	2,67	3,94	2,71	3,51	1,52	1,57	1,31	1,04	1,77	0,84	1,08	2,42	1,03	2,47	4,39	4,49	1,31	1,86	2,74	2,92	2,16	1,62	2,10	2,88	3,75	2,35	2,04
4	1,78	1,91	1,66	0,00	1,40	1,12	2,37	2,95	2,93	1,28	3,53	2,95	3,59	2,19	4,33	1,88	2,27	0,92	2,13	2,80	2,11	0,95	2,16	1,83	3,27	3,93	4,29	1,63	2,72	3,03	3,58	2,54	2,27	1,91	3,36	3,82	2,72	2,94
5	1,07	1,70	1,22	1,40	0,00	1,49	1,10	2,28	1,80	1,11	2,94	2,04	2,87	1,53	4,28	2,23	2,59	1,47	1,81	2,87	1,67	1,05	1,43	1,24	3,39	3,24	3,52	1,01	1,95	3,49	3,90	2,72	2,38	2,14	3,51	4,49	2,51	2,74
6	1,34	1,37	1,00	1,12	1,49	0,00	2,31	3,61	2,49	1,00	3,64	2,84	4,03	2,75	3,45	0,94	1,3	4,063	1,13	1,72	1,22	0,81	2,43	1,18	2,30	4,54	4,67	1,42	2,26	2,22	2,67	1,82	1,27	1,58	2,54	3,12	2,20	2,11
7	1,61	2,51	1,74	2,37	1,10	2,31	0,00	2,37	1,80	1,98	3,36	2,42	3,00	1,72	4,21	2,88	3,07	2,42	2,37	3,31	2,22	2,01	2,04	1,61	3,66	3,34	3,67	1,82	2,35	3,97	4,25	2,94	2,81	2,54	3,79	4,99	2,56	2,78
8	3,21	3,61	3,47	2,95	2,28	3,61	2,37	0,00	3,25	3,08	3,17	2,75	1,88	1,18	6,27	4,33	4,81	3,41	4,00	5,14	3,86	3,06	1,86	3,30	5,53	1,34	2,51	2,85	3,57	5,49	6,05	4,68	4,52	3,76	5,43	6,45	4,12	4,74
9	1,51	1,71	1,96	2,93	1,80	2,49	1,80	3,25	0,00	1,80	2,33	1,36	2,92	2,60	4,50	3,04	3,16	2,43	1,93	3,05	1,72	2,09	1,91	2,07	3,81	3,66	3,25	1,40	0,90	4,19	4,39	3,46	2,96	3,29	4,12	5,17	3,37	3,30
10	1,02	0,85	1,20	1,28	1,11	1,00	1,98	3,08	1,80	0,00	2,67	2,01	3,15	2,17	4,07	1,86	2,09	0,66	1,36	2,35	1,27	0,40	1,68	1,48	3,10	3,82	3,76	0,55	1,55	3,12	3,54	2,51	2,03	2,18	3,36	3,98	2,77	2,79
11	3,05	2,58	3,58	3,53	2,94	3,64	3,36	3,17	2,33	2,67	0,00	1,39	1,79	2,77	6,34	4,38	4,64	3,17	3,52	4,67	3,36	2,95	1,86	3,71	5,54	2,83	1,76	2,31	2,17	5,59	6,03	4,98	4,51	4,50	5,73	6,40	4,98	5,18
12	2,22	1,90	2,67	2,95	2,04	2,84	2,42	2,75	1,36	2,01	1,39	0,00	2,06	2,29	5,38	3,45	3,87	2,57	2,53	3,81	2,40	2,30	1,12	2,60	4,55	2,84	2,31	1,55	1,46	4,66	5,10	4,10	3,55	3,56	4,66	5,61	3,85	4,08
13	3,50	3,48	3,94	3,59	2,87	4,03	3,00	1,88	2,92	3,15	1,79	2,06	0,00	1,83	6,64	4,81	5,18	3,65	4,18	5,40	4,04	3,33	1,79	3,84	5,99	1,37	1,12	2,85	3,18	6,01	6,55	5,15	4,92	4,42	6,00	6,88	4,85	5,35
14	2,48	2,83	2,71	2,19	1,53	2,75	1,72	1,18	2,60	2,17	2,77	2,29	1,83	0,00	5,31	3,50	3,90	2,53	3,20	4,28	3,11	2,20	1,38	2,52	4,67	2,14	2,75	2,06	2,91	4,65	5,20	3,71	3,62	2,90	4,62	5,54	3,35	3,93
15	3,98	4,25	3,51	4,33	4,28	3,45	4,21	6,27	4,50	4,07	6,34	5,38	6,64	5,31	0,00	2,86	2,56	3,97	3,29	2,56	3,54	4,08	5,33	3,21	1,61	7,30	7,40	4,39	4,66	2,23	1,87	1,85	2,28	2,95	1,83	2,51	2,51	1,79
16	2,02	1,98	1,52	1,88	2,23	0,94	2,88	4,33	3,04	1,86	4,38	3,45	4,81	3,50	2,86	0,00	1,11	1,50	1,24	1,29	1,52	1,74	3,15	1,38	1,51	5,30	5,45	2,21	2,85	1,34	1,87	1,44	0,66	1,43	1,70	2,36	1,85	1,58
17	2,05	2,20	1,57	2,27	2,59	1,34	3,07	4,81	3,16	2,09	4,64	3,87	5,18	3,90	2,56	1,11	0,00	1,79	1,55	0,87	1,68	1,97	3,67	1,91	1,45	5,77	5,81	2,48	2,92	1,68	1,65	1,44	1,01	2,11	2,12	2,41	2,43	1,72
18	1,38	1,12	1,31	0,92	1,47	0,63	2,42	3,41	2,43	0,66	3,17	2,57	3,65	2,53	3,97	1,50	1,79	0,00	1,44	2,16	1,45	0,48	2,12	1,60	2,87	4,22	4,27	1,14	2,08	2,73	3,21	2,31	1,82	1,98	3,11	3,53	2,72	2,71
19	1,24	1,14	1,04	2,13	1,81	1,13	2,37	4,00	1,93	1,36	3,52	2,53	4,18	3,20	3,29	1,24	1,55	1,44	0,00	1,32	0,41	1,45	2,56	1,16	2,13	4,78	4,68	1,46	1,71	2,38	2,62	2,14	1,29	2,15	2,51	3,39	2,35	1,99
20	2,18	2,17	1,77	2,80	2,87	1,72	3,31	5,14	3,05	2,35	4,67	3,81	5,40	4,28	2,56	1,29	0,87	2,16	1,32	0,00	1,50	2,33	3,82	2,04	1,28	6,01	5,92	2,61	2,77	1,74	1,54	1,89	1,13	2,51	2,08	2,55	2,62	1,78
21	0,92	1,01	0,84	2,11	1,67	1,22	2,22	3,86	1,72	1,27	3,36	2,40	4,04	3,11	3,54	1,52	1,68	1,45	0,41	1,50	0,00	1,34	2,45	1,26	2,41	4,61	4,48	1,28	1,44	2,71	2,88	2,39	1,60	2,39	2,85	3,71	2,59	2,21
22	1,01	1,06	1,08	0,95	1,05	0,81	2,01	3,06	2,09	0,40	2,95	2,30	3,33	2,20	4,08	1,74	1,97	0,48	1,45	2,33	1,34	0,00	1,83	1,46	3,05	3,88	3,94	0,79	1,82	3,03	3,45	2,46	2,00	2,09	3,31	3,89	2,73	2,74
23	2,06	1,96	2,42	2,16	1,43	2,43	2,04	1,86	1,91	1,68	1,86	1,12	1,79	1,38	5,33	3,15	3,67	2,12	2,56	3,82	1,45	1,83	0,00	2,29	4,43	2,30	2,35	1,34	1,99	4,40	4,95	3,79	3,36	3,02	4,45	5,35	3,47	3,88
24	1,37	1,82	1,03	1,83	1,24	1,18	1,61	3,30	2,07	1,48	3,71	2,60	3,84	2,52	3,21	1,38	1,91	1,60	1,16	2,04	1,26	1,46	2,29	0,00	2,28	4,29	4,51	1,59	2,25	2,47	2,87	1,81	1,35	1,40	2,35	3,55	1,50	1,61
25	3,01	3,15	2,47	3,27	3,39	2,30	3,66	5,53	3,81	3,10	5,54	4,55	5,99	4,67	1,61	1,51	1,45	2,87	2,13	1,28	2,41	3,05	4,43	2,28	0,00	6,55	6,66	3,42	3,76	0,99	0,71	1,39	1,12	2,21	0,98	1,84	2,03	1,12
26	4,00	4,17	4,39	3,93	3,24	4,54	3,34	1,34	3,66	3,82	2,83	2,84	1,37	2,14	7,30	5,30	5,77	4,22	4,78	6,01	4,61	3,88	2,30	4,29	6,55	0,00	1,53	3,49	3,93	6,52	7,08	5,75	5,50	4,87	6,49	7,47	5,23	5,82
27	3,96	3,88	4,49	4,29	3,52	4,67	3,67	2,51	3,25	3,76	1,76	2,31	1,12	2,75	7,40	5,45	5,81	4,27	4,68	5,92	4,48	3,94	2,35	4,51	6,66	1,53	0,00	3,35	3,40	6,71	7,20	5,97	5,61	5,28	6,73	7,62	5,66	6,07
28	0,91	0,81	1,31	1,63	1,01	1,42	1,82	2,85	1,40	0,55	2,31	1,55	2,85	2,06	4,39	2,21	2,48	1,14	1,46	2,61	1,28	0,79	1,34	1,59	3,42	3,49	3,35	0,00	1,15	3,51	3,89	2,91	2,38	2,53	3,67	4,43	3,00	3,04
29	1,35	1,11	1,86	2,72	1,95	2,26	2,35	3,57	0,90	1,55	2,17	1,46	3,18	2,91	4,66	2,85	2,92	2,08	1,71	2,77	1,44	1,82	1,99	2,25	3,76	3,93	3,40	1,15	0,00	4,05	4,24	3,54	2,90	3,41	4,17	4,97	3,67	3,48
30	3,29	3,22	2,74	3,03	3,49	2,22	3,97	5,49	4,19	3,12	5,59	4,66	6,01	4,65	2,23	1,34	1,68	2,73	2,38	1,74	2,71	3,03	4,40	2,47	0,99	6,52	6,71	3,51	4,05	0,00	1,02	1,50	1,27	1,96	0,91	1,18	2,10	1,69
31	3,47	3,55	2,92	3,58	3,90	2,67	4,25	6,05	4,39	3,54	6,03	5,10	6,55	5,20	1,87	1,87	1,65	3,21	2,62	1,54	2,88	3,45	4,95	2,87	0,71	7,08	7,20	3,89	4,24	1,02	0,00	1,84	1,65	2,69	1,37	1,48	2,62	1,76
32	2,66	2,90	2,16	2,54	2,72	1,82	2,94	4,68	3,46	2,51	4,98	4,10	5,15	3,71	1,85	1,44	1,44	2,31	2,14	1,89	2,39	2,46	3,79	1,81	1,39	5,75	5,97	2,91	3,54	1,50	1,84	0,00	0,98	1,27	1,46	2,22	1,41	1,14
33	2,12	2,19	1,62	2,27	2,38	1,27	2,81	4,52	2,96	2,03	4,51	3,55	4,92	3,62	2,28	0,66	1,01	1,82	1,29	1,13	1,60	2,00	3,36	1,35	1,12	5,50	5,61	2,38	2,90	1,27	1,65	0,98	0,00	1,41	1,41	2,28	1,60	1,12
34	2,53	2,70	2,10	1,91	2,14	1,58	2,54	3,76	3,29	2,18	4,50	3,56	4,42	2,90	2,95	1,43	2,11	1,98	2,15	2,51	2,39	2,09	3,02	1,40	2,21	4,87	5,28	2,53	3,41	1,96	2,							

Lampiran 4. Penggabungan *Cluster Hierarki Complete Linkage*

Penggabungan 1

$d_{euclidean}$	(10, 22)	1	2	3	4	5	...	38
(10, 22)	0,00	1,02	1,06	1,20	1,28	1,11	...	2,79
1	1,02	0,00	1,19	0,62	1,78	1,07	...	2,58
2	1,06	1,19	0,00	1,44	1,91	1,70	...	3,03
3	1,20	0,62	1,44	0,00	1,66	1,22	...	2,04
4	1,28	1,78	1,91	1,66	0,00	1,40	...	2,94
5	1,11	1,07	1,70	1,22	1,40	0,00	...	2,74
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	2,79	2,58	3,03	2,04	2,94	2,74	...	0,00

Penggabungan 2

$d_{euclidean}$	(19, 21)	(10, 22)	1	2	3	4	...	38
(19, 21)	0,00	1,45	1,24	1,14	1,04	2,13	...	2,21
(10, 22)	1,45	0,00	1,02	1,06	1,20	1,28	...	2,79
1	1,24	1,02	0,00	1,19	0,62	1,78	...	2,58
2	1,14	1,06	1,19	0,00	1,44	1,91	...	3,03
3	1,04	1,20	0,62	1,44	0,00	1,66	...	2,04
4	2,13	1,28	1,78	1,91	1,66	0,00	...	2,94
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	2,21	2,79	2,58	3,03	2,04	2,94	...	0,00

Penggabungan 3

$d_{euclidean}$	(1, 3)	(19, 21)	(10, 22)	2	4	5	...	38
(1, 3)	0,00	1,24	1,20	1,44	1,78	1,22	...	2,58
(19, 21)	1,24	0,00	1,45	1,14	2,13	1,81	...	2,21
(10, 22)	1,20	1,45	0,00	1,06	1,28	1,11	...	2,79
2	1,44	1,14	1,06	0,00	1,91	1,70	...	3,03
4	1,78	2,13	1,28	1,91	0,00	1,40	...	2,94
5	1,22	1,81	1,11	1,70	1,40	0,00	...	2,74
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	2,58	2,21	2,79	3,03	2,94	2,74	...	0,00

Penggabungan 4

$d_{euclidean}$	(6, 18)	(1, 3)	(19, 21)	(10, 22)	2	4	...	38
(6, 18)	0,00	1,38	1,45	0,81	1,37	1,12	...	2,71
(1, 3)	1,38	0,00	1,24	1,20	1,44	1,78	...	2,58
(19, 21)	1,45	1,24	0,00	1,45	1,14	2,13	...	2,21
(10, 22)	0,81	1,20	1,45	0,00	1,06	1,28	...	2,79
2	1,37	1,44	1,14	1,06	0,00	1,91	...	3,03
4	1,12	1,78	2,13	1,28	1,91	0,00	...	2,94
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	2,71	2,58	2,21	2,79	3,03	2,94	...	0,00

38	1,78	3,04	1,76	1,58	2,71	2,58	...	0,00
----	------	------	------	------	------	------	-----	------

Penggabungan 9

$d_{euclidean}$	(9, 29)	(17, 20)	(10, 22, 28)	(25, 31)	(16, 33)	(6, 18)	...	38
(9, 29)	0,00	3,16	2,09	3,81	3,04	2,49	...	3,48
(17, 20)	3,16	0,00	2,61	1,65	1,29	2,16	...	1,78
(10, 22, 28)	2,09	2,61	0,00	3,89	2,38	1,42	...	3,04
(25, 31)	3,81	1,65	3,89	0,00	1,87	3,21	...	1,76
(16, 33)	3,04	1,29	2,38	1,87	0,00	1,82	...	1,58
(6, 18)	2,49	2,16	1,42	3,21	1,82	0,00	...	2,71
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	3,48	1,78	3,04	1,76	1,58	2,71	...	0,00

Penggabungan 10

$d_{euclidean}$	(30, 35)	(9, 29)	(17, 20)	(10, 22, 28)	(25, 31)	(16, 33)	...	38
(30, 35)	0,00	4,19	4,17	3,67	1,37	1,70	...	1,69
(9, 29)	4,19	0,00	3,16	2,09	3,81	3,04	...	3,48
(17, 20)	4,17	3,16	0,00	2,61	1,65	1,29	...	1,78
(10, 22, 28)	3,67	2,09	2,61	0,00	3,89	2,38	...	3,04
(25, 31)	1,37	3,81	1,65	3,89	0,00	1,87	...	1,76
(16, 33)	1,70	3,04	1,29	2,38	1,87	0,00	...	1,58
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	1,69	3,48	1,78	3,04	1,76	1,58	...	0,00

Penggabungan 11

$d_{euclidean}$	(34, 37)	(30, 35)	(9, 29)	(17, 20)	(10, 22, 28)	(25, 31)	...	38
(34, 37)	0,00	2,10	3,67	2,62	3,00	2,69	...	1,69
(30, 35)	2,10	0,00	4,19	4,17	3,67	1,37	...	1,69
(9, 29)	3,67	4,19	0,00	3,16	2,09	3,81	...	3,48
(17, 20)	2,62	4,17	3,16	0,00	2,61	1,65	...	1,78
(10, 22, 28)	3,00	3,67	2,09	2,61	0,00	3,89	...	3,04
(25, 31)	2,69	1,37	3,81	1,65	3,89	0,00	...	1,76
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	1,69	1,69	3,48	1,78	3,04	1,76	...	0,00

Penggabungan 12

$d_{euclidean}$	(2, 10, 22, 28)	(34, 37)	(30, 35)	(9, 29)	(17, 20)	...	38
(2, 10, 22, 28)	0,00	3,18	3,67	2,09	2,48	...	3,04

(34, 37)	3,18	0,00	2,10	3,67	2,62	...	1,69
(30, 35)	3,67	2,10	0,00	4,19	4,17	...	1,69
(9, 29)	2,09	3,67	4,19	0,00	3,16	...	3,48
(17, 20)	2,48	2,62	4,17	3,16	0,00	...	1,78
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	3,04	1,69	1,69	3,48	1,78	...	0,00

Penggabungan 13

$d_{euclidean}$	(5, 7)	(2, 10, 22, 28)	(34, 37)	(30, 35)	(9, 29)	(17, 20)	...	38
(5, 7)	0,00	2,51	2,56	3,97	2,35	3,31	...	2,78
(2, 10, 22, 28)	2,51	0,00	3,18	3,67	2,09	2,48	...	3,04
(34, 37)	2,56	3,18	0,00	2,10	3,67	2,62	...	1,69
(30, 35)	3,97	3,67	2,10	0,00	4,19	4,17	...	1,69
(9, 29)	2,35	2,09	3,67	4,19	0,00	3,16	...	3,48
(17, 20)	3,31	2,48	2,62	4,17	3,16	0,00	...	1,78
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	2,78	3,04	1,69	1,69	3,48	1,78	...	0,00

⋮

⋮

Penggabungan 35

$d_{euclidean}$	(5, 7, 8, 9, 11, 12, 13, 14, 23, 26, 27, 29)	(15, 16, 17, 20, 25, 30, 31, 32, 33, 34, 35, 36, 37, 38)	(1, 2, 3, 4, 6, 10, 18, 19, 21, 22, 24, 28)
(5, 7, 8, 9, 11, 12, 13, 14, 23, 26, 27, 29)	0,00	7,62	4,49
(15, 16, 17, 20, 25, 30, 31, 32, 33, 34, 35, 36, 37, 38)	7,62	0,00	4,43

(1, 2, 3, 4, 6, 10, 18 19, 21, 22, 24, 28)	4,49	4,43	0,00
---	------	------	------

Penggabungan 36

$d_{euclidean}$	(1, 2, 3, 4, 6, 10, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38)	(5, 7, 8, 9, 11, 12, 13, 14, 23, 26, 27, 29)
(1, 2, 3, 4, 6, 10, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38)	0,00	7,62
(5, 7, 8, 9, 11, 12, 13, 14, 23, 26, 27, 29)	7,62	0,00

Penggabungan 37

d_{euclid}	(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38)
(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38)	7,62

Lampiran 5. Script Proses Analisis *Hierarchical Clustering Multiscale Bootstrap*

```

## Menampilkan data di excel
library(readxl)
datasatu = read_excel("DATA.xlsx")

## Deteksi outlier
boxplot(DATA, main="Data Indikator Pendidikan", boxwex=0.1)

## Menguji multikolinieritas
model = lm(X1~., data = )
car::vif(model)
model2 = lm(X2~., data = )
car::vif(model2)
model3 = lm(X3~., data = )
car::vif(model3)
:
:

model6 = lm(X6~., data = )
car::vif(model6)

## Mengeluarkan variabel X5 dikarenakan terjadi multikolinieritas
datasatu2 = read_excel("DATA BARU.xlsx")

## Mengatasi Multikolinieritas
modell = lm(X1~., data = )
car::vif(modell)
modell2 = lm(X2~., data = )
car::vif(modell2)
modell3 = lm(X3~., data = )
car::vif(modell3)
modell4 = lm(X4~., data = )
car::vif(modell4)
modell5 = lm(X6~., data = )
car::vif(modell5)

## Standardisasi
data_standardized = scale(DATA)
datanew = as.data.frame(data_standardized)

## Menghitung jarak euclidean
dist.eucl = dist(datanew, method = "euclidean")
distance = round(as.matrix(dist.eucl)[1:38, 1:38], 2)
distance

```

```
## complete linkage
pend_completelinkage = hclust(dist.eucl, method = "complete")

## Dendogram complete linkage
library(factoextra)
plot(pend_completelinkage, cex = 0.5, main = "Cluster Dendogram Complete
Linkage")

## hasil akhir cluster
fviz_dend(pend_completelinkage, k=2, k_colors = "jco", rect = T, main =
"Complete Linkage Cluster")

## Multiscale Bootstrap
library(pvclust)
set.seed()
multiscale_clust = pvclust(data, method.dist=" ",
                          method.hclust="complete", nboot = )
plot(multiscale_clust, hang = -1, cex = 0.5)
au_values = pvrect(multiscale_clust, alpha = 0.95)
msplot(multiscale_clust, edges=c( ))
cluster = cutree(pend_completelinkage, 2)
cluster
d2 = as.data.frame(DATA)

## karakterisasi cluster
aggregate(d2,by =list(cluster),
          FUN = mean)
```

RIWAYAT HIDUP



Imro'atul Kalamiyah, lahir di Lamongan pada 17 April 2000. Penulis merupakan anak kedua dari empat bersaudara dari pasangan Bapak Asikin dan Ibu Ernalis Ekowati. Penulis telah menempuh pendidikan formal mulai dari TK Al-Ma'arif Brondong dan lulus pada tahun 2006, dilanjutkan menempuh pendidikan sekolah dasar di MI Al-Ma'arif Brondong dan lulus pada tahun 2012. Kemudian penulis melanjutkan pendidikan sekolah menengah pertama di MTs. Al-Ma'arif Brondong dan lulus pada tahun 2015. Selanjutnya penulis melanjutkan pendidikan menengah atas di MAN 1 Tuban dan lulus pada tahun 2018. Kemudian penulis melanjutkan pendidikan D2 di PROBIS Universitas Negeri Malang sampai tahun 2019 program studi *Computer, English and Banking* tidak sampai lulus, dikarenakan penulis ingin melanjutkan pendidikan Matematika di tempat lain. Oleh karena itu, pada tahun yang sama penulis memutuskan berhenti belajar di PROBIS Universitas Negeri Malang dan melanjutkan pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang program studi Matematika, Fakultas Sains dan Teknologi.

Selama menempuh pendidikan di perguruan tinggi, penulis aktif mengikuti beberapa kegiatan baik di dalam kampus maupun di luar kampus. Penulis bergabung dalam organisasi LBB AHAF Institute di bawah naungan PP. Al-Hikmah Al-Fathimiyyah Malang sebagai pengurus organisasi dan tutor. Penulis juga bergabung dalam kepanitiaan Kompetisi Matematika (KOMET) selama dua periode. Selain itu, penulis juga mengikuti berbagai kegiatan sebagai panitia lomba dan seminar di dalam maupun di luar kampus serta mengikuti berbagai kegiatan di luar kampus seperti pelatihan dan seminar.



BUKTI KONSULTASI SKRIPSI

Nama : Imro'atul Kalamiyah
NIM : 19610110
Fakultas / Program Studi : Sains dan Teknologi / Matematika
Judul Skripsi : Penerapan *Hierarchical Clusterong Multiscale Bootstrap* pada Indikator Pendidikan Tingkat SMA Sederajat di Provinsi Jawa Timur
Pembimbing I : Dr. Fachrur Rozi, M.Si.
Pembimbing II : Ach. Nashichuddin, M.A.

No	Tanggal	Hal	Tanda Tangan
1.	16 Januari 2023	Konsultasi BAB I	1.
2.	18 Januari 2023	Revisi BAB I	2.
3.	08 Februari 2023	Konsultasi BAB II dan III	3.
4.	14 Februari 2023	Revisi BAB II dan III	4.
5.	17 Februari 2023	Konsultasi Kajian Agama	5.
6.	20 Februari 2023	Revisi Kajian Agama	6.
7.	22 Februari 2023	Revisi Kajian Agama	7.
8.	27 Maret 2023	Konsultasi Revisi BAB I, II, dan III	8.
9.	31 Maret 2023	ACC untuk Seminar Proposal	9.
10.	10 Mei 2023	Konsultasi BAB IV	10.
11.	30 Mei 2023	Revisi BAB IV	11.
12.	06 Juni 2023	Konsultasi BAB IV dan V	12.
13.	08 Juni 2023	Revisi BAB IV dan V	13.
14.	08 Juni 2023	Konsultasi Kajian Agama	14.
15.	12 Juni 2023	Revisi Kajian Agama	15.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16.	14 Juni 2023	Revisi Kajian Agama	16.
17.	15 Juni 2023	ACC Seminar Hasil	17.
18.	12 Oktober 2023	Konsultasi Revisi Seminar Hasil	18.
19.	18 Oktober 2023	Konsultasi Revisi Seminar Hasil	19.
20.	01 November 2023	ACC Sidang Skripsi	20.
21.	03 November 2023	ACC Keagamaan	21.
22.	06 Desember 2023	ACC Keseluruhan	22.

Malang, 6 Desember 2023

Mengetahui,

Ka. Program Studi Matematika



Dr. Riky Susanti, M.Sc.

NIP. 19741129 200012 2 005