

**ANALISIS PENGELOMPOKAN NEGARA DI DUNIA
BERDASARKAN INDEKS KEBAHAGIAAN MENGGUNAKAN
METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING**

**M.Fauzan Hidayat¹, Rendy Zikriansyah F.S², Azura Calista³, Arnita⁴, Joel Arie Putranta⁵,
Khaira Nazla⁶**

Universitas Negeri Medan

E-mail: fauzanhidayat773@gmail.com¹, rendy.saragih99@gmail.com², azuracalista9@gmail.com³,
arnita@unimed.ac.id⁴, joelginting34@gmail.com⁵, nazlakhaira911@gmail.com⁶

Abstrak

Penelitian ini bertujuan untuk mengelompokkan negara-negara di dunia berdasarkan tingkat kebahagiaan menggunakan metode Agglomerative Hierarchical Clustering (AHC). Data yang digunakan diambil dari World Happiness Report tahun 2019 dengan variabel seperti GDP per kapita, harapan hidup, kebebasan memilih, kepercayaan, dan kemurahan hati. Setelah dilakukan preprocessing termasuk normalisasi data dan uji multikolinieritas, metode AHC diterapkan menggunakan pendekatan average linkage dan jarak Mahalanobis. Penelitian ini menemukan tiga kluster negara berdasarkan tingkat kebahagiaan: kluster dengan kebahagiaan rendah, sedang, dan tinggi. Hasil visualisasi dendrogram membantu memahami hierarki hubungan antar-negara dalam setiap kluster. Penelitian ini menunjukkan pentingnya faktor-faktor ekonomi dan sosial dalam menentukan tingkat kebahagiaan suatu negara. Purpose: Tujuan dari penelitian ini adalah untuk mengelompokkan negara-negara berdasarkan tingkat kebahagiaan mereka menggunakan metode Agglomerative Hierarchical Clustering (AHC). Motivasi dari penelitian ini adalah untuk mengidentifikasi dan memahami bagaimana faktor-faktor sosial ekonomi dan sosial lainnya, seperti GDP per kapita, harapan hidup, kebebasan memilih, kepercayaan, dan kemurahan hati, memengaruhi kebahagiaan suatu populasi. Dengan membangun dari laporan-laporan sebelumnya tentang kebahagiaan global, penelitian ini bertujuan untuk memberikan pemahaman yang lebih jelas dan berbasis data tentang bagaimana faktor-faktor tersebut saling berinteraksi dan berkontribusi terhadap kebahagiaan keseluruhan di berbagai negara. Temuan utama dari penelitian ini akan membantu para pembuat kebijakan dalam menentukan area yang perlu diperbaiki untuk meningkatkan tingkat kebahagiaan warga negara mereka. Methods/Study design/approach: Penelitian ini menggunakan metode Agglomerative Hierarchical Clustering (AHC) untuk mengklasifikasikan negara-negara berdasarkan tingkat kebahagiaan mereka. Data yang digunakan berasal dari World Happiness Report tahun 2019 dan mencakup beberapa variabel kunci: GDP per kapita, harapan hidup, kebebasan memilih, kepercayaan (diukur melalui persepsi terhadap korupsi), dan kemurahan hati. Proses pra-pengolahan data meliputi deteksi outlier, normalisasi, dan uji multikolinieritas untuk memastikan akurasi dalam proses pengelompokan. Metode average linkage dan jarak Mahalanobis digunakan untuk mengelompokkan negara-negara ke dalam kluster. Teknik visualisasi, seperti dendrogram, digunakan untuk membantu memahami hubungan antar-negara di setiap kluster. Result/Findings: Penelitian ini berhasil mengelompokkan negara-negara di dunia berdasarkan tingkat kebahagiaan menggunakan metode Agglomerative Hierarchical Clustering (AHC) dengan pendekatan Average. Tiga kluster utama terbentuk, yaitu kluster dengan tingkat kebahagiaan rendah, sedang, dan tinggi. Variabel-variabel seperti GDP per kapita, harapan hidup, dukungan sosial, kebebasan, dan kepercayaan memiliki pengaruh signifikan dalam membedakan kluster. Kluster dengan tingkat kebahagiaan tertinggi menunjukkan rata-rata GDP dan harapan hidup yang lebih tinggi dibandingkan kluster lainnya. Penelitian ini juga menemukan adanya korelasi yang kuat antara GDP dan harapan hidup, yang menjadi faktor utama dalam

menentukan tingkat kebahagiaan suatu negara. Dengan perhitungan jarak Mahalanobis dan penggunaan metode Average, kluster yang dihasilkan lebih kompak dan minim varian.. Novelty/Originality/Value: Keunikan dari penelitian ini terletak pada penggunaan metode Agglomerative Hierarchical Clustering (AHC) dengan pendekatan Average dalam mengelompokkan negara berdasarkan indeks kebahagiaan. Penggunaan teknik Principal Component Analysis (PCA) untuk mengatasi masalah multikolinieritas juga merupakan pendekatan baru yang memperbaiki akurasi klusterisasi. Selain itu, penelitian ini memberikan kontribusi dengan menyajikan visualisasi dendrogram yang mempermudah interpretasi hubungan antar-negara berdasarkan faktor-faktor kebahagiaan. Penelitian ini memberikan nilai tambah dengan analisis mendalam terhadap data kebahagiaan global dan mampu mengidentifikasi faktor-faktor utama yang berkontribusi pada kebahagiaan di berbagai negara, yang dapat digunakan sebagai dasar pengambilan kebijakan untuk meningkatkan kesejahteraan global.

Kata Kunci — Writing Guidelines, Scientific Manuscripts, JIDS.

1. PENDAHULUAN

Kebahagiaan dipandang sebagai salah satu elemen penting dalam kehidupan masyarakat. Tidak hanya bertujuan untuk kebahagiaan pribadi, namun banyak yang berpendapat bahwa kebahagiaan dapat diciptakan tidak hanya dari diri sendiri namun juga pengaruh orang lain, bahkan pemerintah bertanggung jawab untuk meningkatkan kebahagiaan Masyarakat secara keseluruhan.

Kebahagiaan suatu negara dapat diukur dari kesejahteraan warganya. Negara yang memiliki kesejahteraan tinggi cenderung memiliki Masyarakat yang lebih bahagia karena kebutuhan mereka terpenuhi, yang pada akhirnya meningkatkan peluang mereka untuk merasa bahagia. Konsep kesejahteraan dalam negara modern berakar pada keyakinan bahwa kebahagiaan individu dapat ditingkatkan dengan memperbaiki kondisi kehidupan.[1]

Kebahagiaan terkadang dianggap penting oleh Masyarakat, oleh karena itu ada banyak penelitian yang telah dilakukan mengenai kebahagiaan. Perserikatan Bangsa-Bangsa (PBB) juga melakukan penelitian terkait dengan tingkat kebahagiaan di dunia, seperti penerbitan World Happiness Report pada tahun 2012 dan terus berlanjut setiap tahunnya hingga sekarang. World Happiness Report yang diterbitkan oleh PBB ini dilakukan dengan memberikan pertanyaan subjektif kepada 1.000 orang lebih di 150 lebih negara. Berbagai jenis statistik juga digunakan untuk menjelaskan kenapa sebuah negara lebih bahagia dibandingkan dengan negara yang lainnya. Ada beberapa faktor yang dikaji yaitu kekuatan ekonomi berdasarkan GDP per kapita, dukungan social, tingkat harapan hidup, kebebasan untuk memilih, kemurahan hati, dan juga persepsi terkait dengan korupsi.[1]

Banyak penelitian yang telah dilakukan untuk menganalisis faktor-faktor yang mempengaruhi tingkat kebahagiaan di suatu negara dan melakukan pengelompokan berdasarkan indeks kebahagiaan tersebut.

Salah satunya adalah penelitian yang dilakukan oleh [2] yang membandingkan algoritma K-Means dan Fuzzy C-Means untuk mengelompokkan provinsi di Indonesia berdasarkan indeks kebahagiaannya. Selain itu, analisis cluster dapat diterapkan pada berbagai jenis data untuk mengelompokkan objek-objek berdasarkan kemiripan yang dimiliki. Misalnya, penelitian oleh [3] menggunakan algoritma Agglomerative Hierarchical Clustering (AHC) untuk mengelompokkan provinsi di Indonesia berdasarkan tingkat kemiskinan.

Berdasarkan penelitian tersebut, peneliti tertarik untuk melakukan pengelompokan negara-negara di dunia berdasarkan tingkat kebahagiaan menggunakan metode Agglomerative Hierarchical Clustering (AHC). Metode ini dipilih karena kemampuan AHC untuk menghasilkan dendrogram, yang memberikan visualisasi hierarki hubungan antar negara berdasarkan faktor-faktor kebahagiaan. Penelitian ini bertujuan untuk mengetahui kondisi tingkat kebahagiaan di berbagai negara, mengidentifikasi

pengelompokan tingkat kebahagiaan, serta memahami karakteristik tiap kluster yang terbentuk.

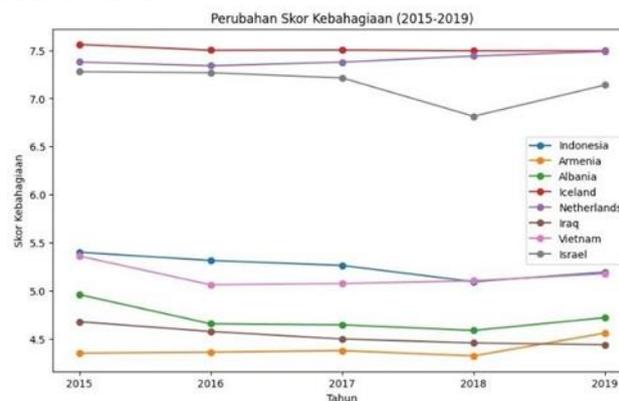
Metode Agglomerative Hierarchical Clustering (AHC) ini dipilih karena kelebihanannya yaitu kemampuannya untuk menangani banyak cluster dan tidak memerlukan penentuan jumlah cluster sebelumnya.[19] Metode Agglomerative Clustering menghasilkan dendrogram, yaitu representasi visual dari penggabungan kluster dalam bentuk pohon. Dendrogram ini berguna untuk memahami struktur hierarki data.[19] Selain itu, metode ini efektif dalam menangani data yang bersifat multivariat dan mampu mengelompokkan objek yang memiliki karakteristik serupa secara bertahap, dimulai dari objek individual hingga membentuk kluster yang lebih besar. Proses ini memungkinkan analisis mendalam terkait struktur dan pola kemiripan antar objek, dalam hal ini negara-negara berdasarkan tingkat kebahagiaan mereka.

2. METODE PENELITIAN

Penelitian ini menggunakan metode hierarchical clustering untuk mengelompokkan negara berdasarkan tingkat kebahagiaan. Hierarchical clustering merupakan pengembangan dari suatu hirarki dengan struktur mirip phon bercabang. Pengelompokan hirarki memiliki algoritma dengan melakukan pembuatan m cluster yang terdiri atas 1 objek dari suatu data. Kemudian antara satu cluster dengan cluster lain yang berdekatan akan digabung, tahapan tersebut akan terus berlanjut hingga m cluster yang terdiri atas satu objek menjadi satu cluster dengan anggota m objek[3].

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan metode Hierarchical Clustering dengan algoritma agglomerative untuk mengelompokkan negara-negara di dunia berdasarkan tingkat kebahagiaannya. Langkah pertama yang dilakukan dalam proses pengelompokan ini adalah analisis deskriptif. Berikut ini akan disajikan gambaran tingkat kebahagiaan beberapa di dunia dari 2015-2019.



Gambar 1: Visualisasi Skor Kebahagiaan Negara di Dunia 2015-2019

Berdasarkan Gambar 1 skor kebahagiaan beberapa negara di dunia cenderung mengalami masa naik dan turun. Bahkan ada beberapa negara di wilayah yang sama dapat berbanding terbalik kebahagiaan masyarakatnya seperti iran dan israel.

Selanjutnya akan dilakukan eksplorasi dataset, termasuk mengeksplorasi variabel-variabel yang ada. Dalam dataset yang digunakan ini terdapat 6 variabel independent seperti GDP, Family Support, Life Expectancy (Health), Freedom, Trust, Generosity untuk melihat skor kebahagiaan dan melakukan perankingan negara berdasarkan tingkat score atau tingkat kebahagiaannya. Secara umum dapat dilihat pada Tabel 1 berikut.

Table 1. Data World Happiness Report 2019

No	Country	Region	Rank 2019	Score 2019	GDP 2019	Family 2019	Life Expectancy 2019	Freedom 2019	Trust 2019	Generosity 2019
0	Afganistan	Southern Asia	154	3.203	0.350	0.517	0.361	0.000	0.025	0.158
1	Albania	Central and Eastern Europe	107	4.719	0.947	0.874	0.874	0.383	0.027	0.178
2	Algeria	Middle East and Northern Africa	88	5.211	1.002	1.160	0.785	0.086	0.114	0.073

Kemudian kita akan melakukan analisis deskriptif dengan melihat nilai minimum, maksimum, mean, standar deviasi, kuartil 1 dan juga kuartil 3. Setelah melihat hal tersebut kita dapat menyimpulkan secara tersirat apakah dataset kita berdistribusi normal atau tidak, dan terdapat outlier atau tidak yang akan disajikan pada Tabel 2.

Tabel 2. Deskripsi Data

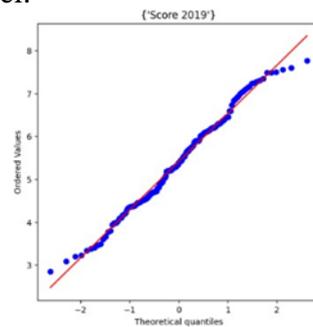
	Rank 2019	Score 2019	GDP 2019	Family 2019	Life Expectancy 2019	Freedom 2019	Trust 2019	Generosity 2019
Count	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000
Mean	78.500000	0.905147	0.905147	1.208814	0.725244	0.392571	0.110603	0.184846
Std	45.177428	0.398389	0.398389	0.299191	0.242124	0.143289	0.094538	0.095254
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	39.750000	0.602750	0.602750	1.055750	0.547750	0.308000	0.047000	0.108750

Tabel 2 menyajikan informasi terkait mean, standar deviasi, nilai minimum dan maksimum, serta nilai kuartil pertama sampai kuartil ketiga dari setiap variabel. Berdasarkan nilai mean, variabel score menunjukkan angka 0.905 yang dapat mengindikasikan rata-rata indeks kebahagiaan di negara tersebut cenderung positif. Kemudian untuk variabel lain seperti GDP, Family, Life Expectancy, Freedom, Trust dan Generosity menunjukkan rata-rata yang cukup bervariasi. GDP dengan nilai mean 0.905 menandakan tingkat kesejahteraan ekonomi memiliki kontribusi yang kuat terhadap kebahagiaan. Family dengan mean 1.209 juga menunjukkan pentingnya faktor social dalam indeks kebahagiaan. Life Expectancy dengan mean 0.725 menunjukkan pentingnya kesehatan dan harapan hidup dalam kebahagiaan seseorang. Variabel yang lain juga berperan namun dengan nilai mean yang lebih kecil seperti Generosity dengan mean 0.185.

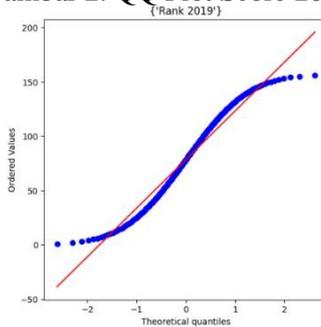
Simpangan baku cukup bervariasi di setiap variabel. Trust memiliki simpangan baku sebesar 0.143 menunjukkan terdapat perbedaan besar dalam tingkat kepercayaan Masyarakat antar negara. Nilai kuartil 1 (25%), kuartil 2 (50%), dan kuartil 3 (75%) dapat memberikan gambaran tentang distribusi data. 50% untuk score ada 0.960 yang berarti setengah dari negara di dunia memiliki skor kebahagiaan dibawah atau sama dengan nilai ini, sedangkan setengahnya lagi diatasnya. Kemudian 75% untuk Freedom adalah 0.882

yang menunjukkan bahwa Sebagian besar negara memiliki tingkat kebebasan yang moderat tinggi. Berdasarkan nilai tersebut kita juga dapat melihat distribusi data, nilai mean dari Score adlah 0.905 dan nilai 50% dari score adalah 0.960. Pada distribusi normal nilai mean dan median harusnya mendekati satu sama lain. Dalam kasus ini, hanya terdapat perbedaan kecil yang menunjukkan kemungkinan berdistribusi normal atau simetris, namun bisa sedikit condong ke kiri (left skewed) karena nilai mean lebih kecil dari median.

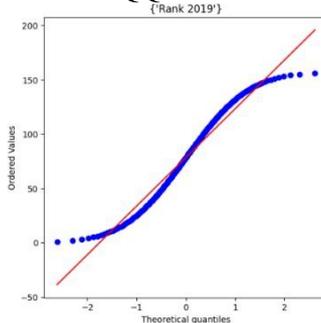
Untuk memastikan apakah data yang akan dianalisis berdistribusi normal atau tidak, dapat dilakukan uji Shapiro wilk untuk melihat distribusi data atau bisa juga melakukan Q-Q Plot. Uji Shapiro wilk dilakukan untuk mengetahui sebaran data acak suatu sampel kecil kurang dari 50 sampel. Suatu data dikatakan berdistribusi normal apabila nilai signifikansi lebih dari 0.05 [17]. Setelah melakukan uji shapiro wilk terdapat beberapa variabel yang tidak berdistribusi normal seperti life expetancy dan freedom. Selain menggunakan uji shapiro wilk, pada penelitian ini akan disajikan juga visualisasi dari QQ plot untuk melihat persebaran data dari setiap variabel. Pada Gambar 2 sampai Gambar 9 akan disajikan visualisasi QQ plot setiap variabel.



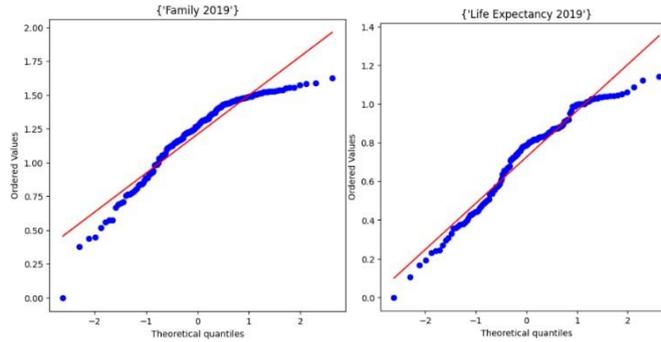
Gambar 2: QQ Plot Score 2019



Gambar 3: QQ Plot Rank 2019

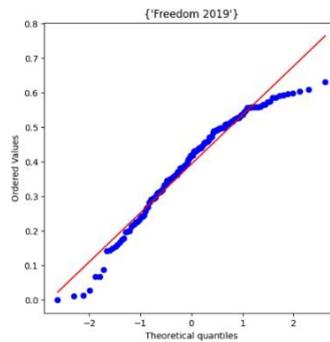


Gambar 4 : QQ Plot GDP 2019

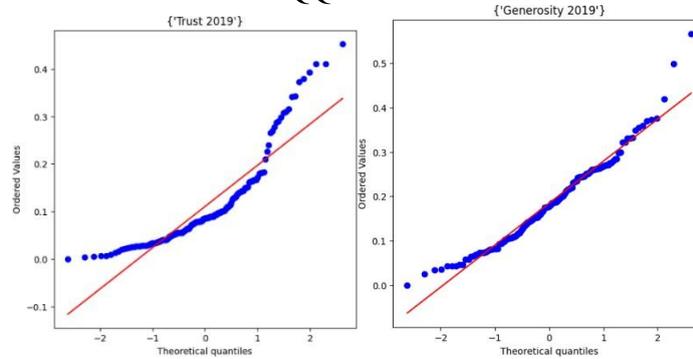


Gambar 5 : QQ Plot Family 2019

Gambar 6 : QQ Plot Life Expetacy 2019



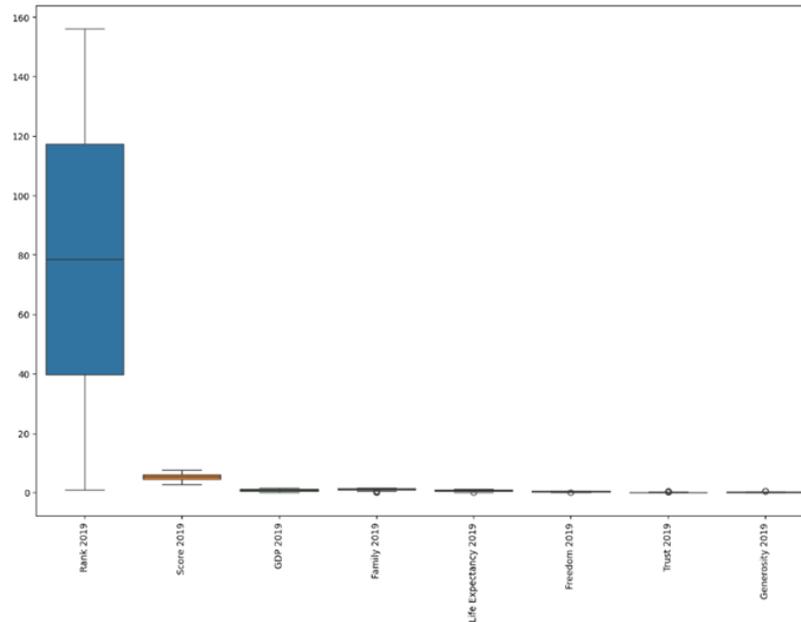
Gambar 7: QQ Plot Freedom 2019



Gambar 8: QQ Plot Trust 2019 Gambar 9: QQ Plot Generosity 2019

Berdasarkan hasil visualisasi QQ Plot setiap variabel dapat dilihat variabel seperti score 2019 menunjukkan data berdistribusi normal namun pada bagian ekor sedikit menyimpang yang menandakan mungkin terdapat outlier. Kemudian variabel lainnya seperti Trust 2019 menunjukkan penyimpangan yang sangat jelas sehingga dapat disimpulkan data tidak berdistribusi normal. Hasilnya sama dengan hasil uji shapiro wilk.

Selanjutnya kita akan melakukan tahapan preprocessing data dimulai dari pengecekan missing value, hingga data reduction. Pada dataset world happiness report 2019 tidak terdapat missing value. Kemudian dilanjutkan dengan memvisualisasikan boxplot untuk melihat apakah terdapat data yang outlier. Pada Gambar 10 akan disajikan visualisasi boxplot.



Gambar 10: Visualisasi Boxplot

Berdasarkan visualisasi dari Gambar 10 terlihat terdapat beberapa outlier di beberapa variabel mulai dari family sampai dengan generosity. Pada penelitian ini juga dilakukan perhitungan interquartile Range (IQR) agar dapat terlihat dengan jelas objek apa yang merupakan data outlier, dalam hal ini adalah negara mana yang merupakan data outlier di setiap variabel. IQR merupakan teknik untuk menemukan data yang merupakan outlier yang terdistribusi. IQR merupakan perbedaan antara kuartil pertama dengan kuartil ketiga [18]. Setelah dilakukan perhitungan IQR ditemukan beberapa data yang outlier. Outlier terbanyak terdapat di variabel trust, terdapat 14 negara yang merupakan outlier dalam variabel ini. Outlier sendiri dapat mengganggu proses analisis namun terkadang juga dapat sangat membantu dalam proses analisis. Dalam hal ini ada kemungkinan terdapat perbedaan tingkat kepercayaan yang cukup signifikan diantara negara-negara di dunia sehingga pada penelitian ini data outlier akan tetap dipertahankan.

Selanjutnya akan ditampilkan visualisasi heatmap untuk melihat korelasi antar variabel. Pada Gambar 11 akan disajikan visualisasi heatmap korelasi variabel.



Gambar 11: Heatmap Korelasi antar Variabel

Berdasarkan Gambar 11 dapat dilihat ada beberapa variabel yang mempengaruhi score 2019 diantaranya seperti GDP, Family, dan Life Expetacy. Ini menandakan variabel tersebut la yang sangat mempengaruhi nilai dari score 2019 sedangkan variabel lainnya hanya memberikan dampak yang tidak begitu besar. Kemudian juga dapat dilihat bahwa ada variabel yang saling mempengaruhi satu lain atau memiliki korelasi yang cukup tinggi seperti GDP dan Life Expetacy yaitu sekitar 0.84. Hal tersebut dapat mengindikasikan terjadinya multikolinieritas diantara variabel tersebut. Untuk memastikan apakah terjadi multikolinieritas kita dapat melihat nilai variance inflating factor (VIF). Jika nilai VIF

lebih dari 10 ada kemungkinan terjadi gejala multikolinieritas. Apabila terjadi multikolinieritas maka akan dilakukan teknik PCA untuk mereduksi data dan mengurangi korelasi antar variabel. Pada Tabel 3 akan menampilkan hasil perhitung VIF dari setiap variabel.

Tabel 3 . Hasil Perhitungan VIF

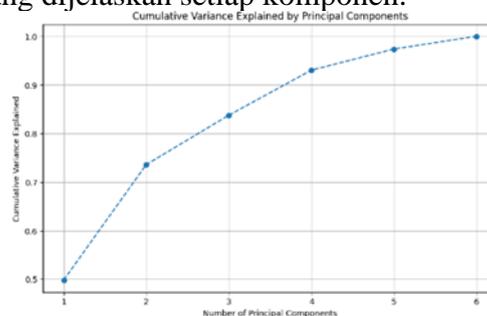
Variabel	VIF
Rank 2019	3.43
Score 2019	77.69
GDP 2019	27.94
Family 2019	55.36
Life Expetacy 2019	38.32
Freedom 2019	14.83
Trust 2019	3.45
Generosity 2019	5.92

Berdasarkan hasil perhitungan VIF terdapat beberapa variabel yang memiliki nilai VIF lebih dari 10 sehingga ada gejala terjadi multikolinieritas. Pada penilitan kali ini akan dilakukan teknik PCA untuk mengatasi adanya multikolinieritas. Untuk melakukan PCA kita harus melakukan standarisasi pada data terlebih dahulu agar skala yang sama. Pada penelitian kali ini standarisasi dilakukan dengan metode Z- Score atau standardscaler(). Setelah dilakukan standarisasi selanjutnya akan dilakukan perhitungan nilai VIF untuk membuktikan apakah PCA dapat menyelesaikan permasalahan multikolinieritas atau tidak. Pada Tabel 4 akan disajikan hasil perhitungan VIF pada data yang telah distandarisasi.

Tabel 4 . Hasil Perhitungan VIF setelah Standarisasi

Variabel	VIF
Rank 2019	9.82
Score 2019	8.21
GDP 2019	4.51
Family 2019	3.05
Life Expetacy 2019	3.85
Freedom 2019	1.72
Trust 2019	1.44
Generosity 2019	1.22

Standarisasi data berhasil mengurangi nilai korelasi antar variabel , walaupun masih terdapat nilai yang cukup tinggi pada Rank 2019. Selanjutnya kita akan melakukan PCA untuk mereduksi dataset sehingga daoat dipastikan tidak ada variabel yang saling berkorelasi. Hasilnya 4 komponen PCA akan digunakan untuk proses pengelompokkan. Dasar pemilihan 4 komponen yang akan digunakan adalah dengan melihat visualisasi varians yang dijelaskan di setiap komponennya. Pada Gambar 12 akan disajikan visualisasi dari varians yang dijelaskan setiap komponen.

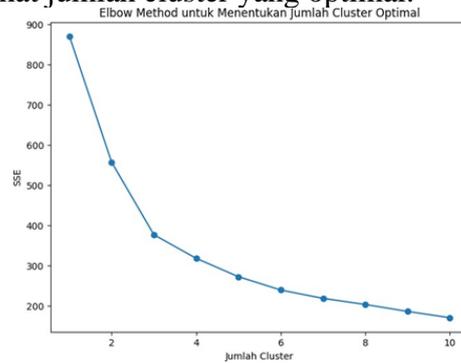


Gambar 12: Visualisasi Varians setiap komponen

Berdasarkan Gambar 12 dapat dilihat pada komponen 4 varians yang dijelaskan

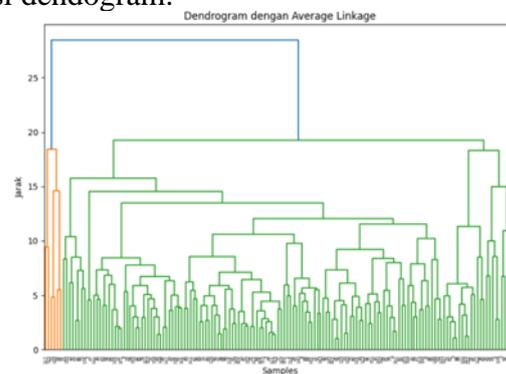
sudah diatas 90% dan untuk komponen selanjutnya tidak menunjukkan kenaikan yang terlalu signifikan sehingga akan digunakan 4 komponen PCA. Optimalnya sedikit komponen namun varians yang dijelaskan tinggi merupakan dasar pemilihan jumlah komponen dari PCA.

Selanjutnya akan ditentukan jumlah cluster yang optimal dengan menggunakan elbow methods atau dengan melihat nilai SSE. Berikut akan ditampilkan visualisasi dari elbow methods untuk melihat jumlah cluster yang optimal.



Gambar 13: Visualisasi Elbow Methods

Pada Gambar 13 dapat dilihat penurunan secara signifikan berada diantara kluster 1 sampai dengan kluster 3 sedangkan selanjutnya penurunan mulai dengan bertahap. Jadi dapat disimpulkan jumlah kluster yang optimal adalah sebanyak 3 kluster. Setelah menentukan jumlah kluster yang optimal kita dapat melakukan pengelompokkan dengan pendekatan AHC. Pada tahapan pengimplementasian algoritma AHC setelah menentukan jumlah kluster yang optimal selanjutnya adalah menghitung jarak setiap kluster, pada penelitian kali ini akan dilakukan perhitungan jarak Mahalanobis. Dalam penelitian ini, hasil perhitungan jarak Mahalanobis antar kluster adalah 2.7366. Hasil ini menunjukkan seberapa jauh dua kluster berdasarkan distribusi data dan variabilitas antar variabel. Selanjutnya adalah melakukan pengelompokkan menggunakan AHC dengan pendekatan average linkage. Setelah dilakukan pengelompokkan dengan jumlah 3 kluster maka dihasilkan visualisasi dendrogram untuk melihat jarak setiap klasternya. Pada Gambar 14 akan disajikan visualisasi dendrogram.



Gambar 14: Visualisasi Dendrogram

Selanjutnya akan dilakukan profilisasi setiap kluster, pertama akan dilihat berapa jumlah negara yang ada disetiap kluster. Pada Tabel 5 akan ditampilkan jumlah negara disetiap kluster yang terbentuk.

Tabel 5. Data jumlah Negara setiap Kluster

Kluster	Jumlah Negara
1	6
2	126
3	24

Kemudian akan dilakukan perhitungan rata-rata pada variabel score untuk melihat klaster mana yang memiliki tingkat kebahagiaan tinggi, sedang, atau rendah. Berikut akan ditampilkan hasil perhitungan rata-rata kebahagiaan di setiap klaster.

Tabel 6. Perhitungan Rata-Rata Kebahagiaan

Klaster	Rata-Rata
1	4.102
2	5.350
3	6.033

Pada Tabel 6 dapat dilihat klaster 1 memiliki rata-rata 4.102, klaster ini memiliki rata-rata yang rendah dibandingkan 2 klaster lainnya. Hal ini menandakan bahwa klaster 1 merupakan klaster dengan tingkat kebahagiaan yang rendah. Kemudian klaster 2 memiliki rata-rata 5.350, klaster ini memiliki nilai rata-rata yang sedang. Hal tersebut menandakan bahwa klaster 2 merupakan klaster dengan tingkat kebahagiaan sedang. Lalu klaster 3 memiliki nilai rata-rata yang cukup tinggi sehingga dapat dikatakan bahwa klaster 3 merupakan klaster dengan tingkat kebahagiaan yang tinggi. Pada Tabel 7 akan ditampilkan hasil profilisasi klaster.

Tabel 7. Profilisasi Klaster

Klaster	Jumlah Negara	Tingkat Kebahagiaan
1	6	Rendah
2	126	Sedang
3	24	Tinggi

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode Agglomerative Hierarchical Clustering (AHC) mampu mengelompokkan negara-negara berdasarkan tingkat kebahagiaannya dengan pendekatan yang komprehensif. Dengan menggunakan variabel-variabel seperti GDP per kapita, harapan hidup, dukungan sosial, dan kebebasan, negara-negara dikelompokkan ke dalam tiga klaster utama: klaster dengan tingkat kebahagiaan rendah, sedang, dan tinggi. Proses analisis memperlihatkan bahwa negara dengan GDP dan harapan hidup yang lebih tinggi cenderung berada dalam klaster dengan tingkat kebahagiaan yang lebih baik.

Penggunaan metode Average dalam AHC terbukti efektif dalam menghasilkan klaster yang kompak dengan varian yang minimal. Selain itu, analisis multikolinieritas yang diatasi dengan teknik PCA memungkinkan penyusunan klaster yang lebih akurat. Hasil visualisasi dendrogram dan analisis karakteristik klaster juga memberikan wawasan mendalam mengenai faktor-faktor yang mempengaruhi kebahagiaan di berbagai negara. Penelitian ini diharapkan dapat menjadi dasar untuk analisis kebijakan yang lebih baik dalam meningkatkan kesejahteraan global.

DAFTAR PUSTAKA

- D. Ayu Kumalasari and I. Gusti Wayan Murjana Yasa, "PENGARUH FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KEBAHAGIAAN NEGARA DI DUNIA," *E-JURNAL EKONOMI PEMBANGUNAN UNIVERSITAS UDAYANA*, vol. 9. No 5, pp. 963–992, 2020.
- Inna Auliya, F. Fitri, Nonong Amalita, and Tessa Octavia Mukhti, "Comparison of K-Means and Fuzzy C-Means Algorithms for Clustering Based on Happiness Index Components Across

- Provinces in Indonesia,” *UNP Journal of Statistics and Data Science*, vol. 2, no. 1, pp. 114–121, Feb. 2024, doi: 10.24036/ujsds/vol2-iss1/150.
- E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, “Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering (Indonesian Province Grouping Based on Poverty Level Using Hierarchical Agglomerative Clustering Analysis),” *Seminar Nasional Official Statistics*, pp. 557–566, 2021.
- E. K. Tokuda, C. H. Comin, and L. da F. Costa, “Revisiting agglomerative clustering,” *Physica A: Statistical Mechanics and its Applications*, vol. 585, Jan. 2022, doi: 10.1016/j.physa.2021.126433.
- E. Sutadji et al., *ANALISIS DATA MULTIVARIAT*. 2022.
- L. Ramos Emmendorfer and A. M. de Paula Canuto, “A generalized average linkage criterion for Hierarchical Agglomerative Clustering,” in *Applied Soft Computing*, Elsevier Ltd, Mar. 2021. doi: 10.1016/j.asoc.2020.106990.
- A. M. Jarman, “Hierarchical Cluster Analysis: Comparison of Single linkage, Complete linkage, Average linkage and Centroid Linkage Method”, doi: 10.13140/RG.2.2.11388.90240.
- D. Exasanti and A. Jananto, “Analisa Hasil Pengelompokan Wilayah Kejadian Non-Kebakaran Menggunakan Agglomerative Hierarchical Clustering pada Dinas Pemadam Kebakaran Kota Semarang,” 2021.
- A. Suarisman, A. Nazir, F. Syafria, and L. Afriyanti, “Perbandingan Jarak Metrik pada Klasifikasi Jamur Beracun Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 10–19, Nov. 2023, doi: 10.47065/josyc.v5i1.4511.
- A. Smiti, “A critical overview of outlier detection methods,” Nov. 01, 2020, Elsevier Ireland Ltd. doi: 10.1016/j.cosrev.2020.100306.
- Effiyaldi et al., “PENERAPAN UJI MULTIKOLINERITAS DALAM PENELITIAN MANAJEMEN SUMBER DAYA MANUSIA,” *Jurnal Ilmiah Manajemen dan Kewirausahaan*, pp. 94–102, 2022, [Online]. Available: <https://ejournal.unama.ac.id/index.php/jumanager>
- F. Mega Kusuma and A. Wibowo, “PRINCIPAL COMPONENT ANALYSIS (PCA) UNTUK MENGATASI MULTIKOLINERITAS TERHADAP FAKTOR ANGKA KEJADIAN PNEUMONIA BALITA DI JAWA TIMUR TAHUN 2014,” *Jurnal Biometrika dan Kependudukan*, vol. Vol. 6, No.2, pp. 89–97, 2017.
- E. Umargono, J. E. Suseno, and V. Gunawan, “K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula,” 2020.
- R. O. Pratikto and N. Damastuti, “Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir,” *Journal of Information Technology and Conmputer Science*, vol. 6, no. 1, pp. 13–20, 2021.
- P. Kumari and S. Gupta, “Comparative analysis between Euclidean distance metric and Mahalanobis Distance Metric,” *International Journal of Innovative Research in Technology and Science* www.ijirts.org, vol. 12, 2024, [Online]. Available: www.ijirts.org
- Y. Asyfani et al., “Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Kepadatan Penduduk Menggunakan Metode Hierarchical Clustering Info Artikel,” vol. 2, no. 1, pp. 1–8, 2024, doi: 10.26714/jodi.
- W. A. Nurcahya, N. P. Arisanti, and A. N. Hanandhika, “Penerapan Uji Asumsi Klasik Untuk Mendeteksi Kesalahan Pada Data Sebagai Upaya Menghindari Pelanggaran Pada Asumsi Klasik,” vol. 1, doi: 10.5281/zenodo.104492725.
- H. P. Vinutha, B. Poornima, and B. M. Sagar, “Detection of outliers using interquartile range technique from intrusion dataset,” in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2018, pp. 511–518. doi: 10.1007/978-981-10-7563-6_53.
- S. I. Attaqwa, A. A. Hanafi, H. Hakim, A. A. Sofyan, and A. P. Sari, “Customer clustering menggunakan K-Means dan Agglomerative pada pendapatan dan pembelian daging,” *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 3, pp. 2747-0563, 2023.