



## **Klasterisasi *World Map Asia* menggunakan Algoritma *Agglomerative Clustering*** (*World Map Asia clustering using Agglomerative Clustering Algorithm*)

Faridah Dewi Khansa<sup>1</sup>, Nisa Eka Juliana<sup>2</sup>, Aaz M Hafidz Azis<sup>3</sup>, Rafli Indra Gunawan<sup>4</sup>

<sup>1</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050042@student.uinsgd.ac.id

<sup>2</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050116@student.uinsgd.ac.id

<sup>3</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050001@student.uinsgd.ac.id

<sup>4</sup>Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050093@student.uinsgd.ac.id

### **Abstrak**

Analisis cluster adalah salah satu teknik yang dapat dikorelasikan. Jelaskan kedekatan atau kesamaan antara objek dan variabel. Analisis terbagi atas 2 metode, yaitu hirarki dan non-hirarki. Dalam metode hirarki terdapat beberapa metode antara lain metode complete linkage, average linkage, single linkage dan ward linkage. Metode ini diterapkan dalam pengelompokan benua Asia. Tujuan dari penelitian ini adalah untuk mengelompokkan negara berdasarkan GDP dan populasi di Asia. Pada penelitian ini metode hirarki yang digunakan adalah Agglomerative Hierarchical Clustering (AHC). 177 negara dengan 6 kolom data negara di seluruh dunia dan di sortir menjadi 47 data negara dan 6 kolom data negara di benua Asia. Pendapatan negara perkapita diperoleh dengan pembagian antara GDP dengan Populasi. Kemudian dibuat 4 kluster. Kluster 1 terdiri atas negara United Arab Emirates, Kuwait, dan Qatar. Kluster 2 terdiri atas Brunei, Oman, Saudi Arabia, dan Taiwan. Kluster 3 terdiri atas 34 negara dan kluster terakhir yaitu Cyprus, Israel, Kazakhstan, South Korea, Malaysia.

**Kata kunci:** agglomerative hierarchical clustering, clustering, data mining

### **Abstract**

*Cluster analysis is one technique that can be correlated. Explain the closeness or similarity between objects and variables. The analysis is divided into 2 methods, namely hierarchical and non-hierarchical. In the hierarchical method, there are several methods, including the complete linkage method, average linkage, single linkage and ward linkage. This method is applied in the grouping of the Asian continent. The purpose of this study is to classify countries based on GDP and population in Asia. In this study, the hierarchical method used is Agglomerative Hierarchical Clustering (AHC). 177 countries with 6 columns of country data around the world and sorted into 47 country data columns and 6 country data columns in the Asian continent. The per capita state income is obtained by dividing GDP by population. Then made 4 clusters. Cluster 1 consists of the United Arab Emirates, Kuwait and Qatar. Cluster 2 consists of Brunei, Oman, Saudi Arabia and Taiwan. Cluster 3 consists of 34 countries and the last cluster is Cyprus, Israel, Kazakhstan, South Korea, Malaysia.*

**Keywords:** agglomerative hierarchical clustering, clustering, data mining

## 1 Pendahuluan

Asia adalah benua terbesar dan terpadat di dunia. Luas wilayahnya kurang lebih 44,58 juta kilometer persegi. Menurut laporan di situs resmi Perserikatan Bangsa-Bangsa, populasi Asia pada tahun 2016 adalah 4,5 miliar, terhitung 60% dari populasi bumi [1]. Saat ini, Asia terdiri dari 48 negara berdaulat. Ada juga tiga wilayah khusus, yaitu Hong Kong dan Makau Cina, dan Taiwan yang disengketakan dengan Cina. Selain dikenal sebagai benua terpadat penduduknya, benua Asia juga merupakan benua terbesar. Budaya, geografi, dan kinerja ekonomi Asia juga berbeda. Ada beberapa negara di Asia yang bisa menjadi negara yang bisa memberikan nafkah yang cukup tinggi kepada warganya. Pendapatan ini berasal dari industri yang cukup maju di sektor sumber daya alam dan sektor lainnya. Negara terkaya di Asia adalah negara dengan pendapatan tinggi. Suatu negara dapat digolongkan sebagai kategori terkaya karena nilai PDB (Produk Domestik Bruto) atau yang biasa disebut dengan PDB (Produk Domestik Bruto).

Clustering adalah metode yang digunakan dalam data mining, bagaimana cara kerjanya dan mengelompokkan data dengan kesamaan fitur antara satu data dengan data lainnya sudah didapat [2]. Pengelompokan data dapat dilakukan dengan konsep data mining meliputi pengumpulan data, pola hubungan dalam kumpulan data ukuran besar dan penggunaan historis dari urutan [3]. Hasil data mining dapat digunakan untuk penentuan keputusan masa depan. Dalam kasus yang dibahas dalam penelitian ini, salah satu teknik yang digunakan dalam data mining adalah clustering. Pengelompokan membagi kumpulan data menjadi beberapa kelompok kemiripan kelompoknya lebih besar dari pada kelompok lain [4]. Sampai saat ini, ada beberapa metode membuat cluster, baik data dokumen maupun untuk data lainnya.

Untuk mengetahui situasi individu atau kelompok, maka dilakukan klasifikasi berdasarkan keadaan tingkat kesejahteraan sosial ekonomi Publik. Pengelompokan sosial ekonomi masyarakat bisa dilakukan dengan metode atau analisis cluster. Pada dasarnya sebuah analisis cluster adalah metode pengelompokan di mana grup memiliki karakteristik yang relatif sama (homogen) saat antar kelompok memiliki karakteristik yang berbeda. Secara umum, objek dimasukkan ke dalam file cluster atau grup, sehingga mereka lebih cenderung terkait (berkorelasi) dengan objek lain di cluster selain dengan objek dari cluster lain. Pembentukan cluster tergantung pada apakah hubungan antar objek kuat atau tidak.

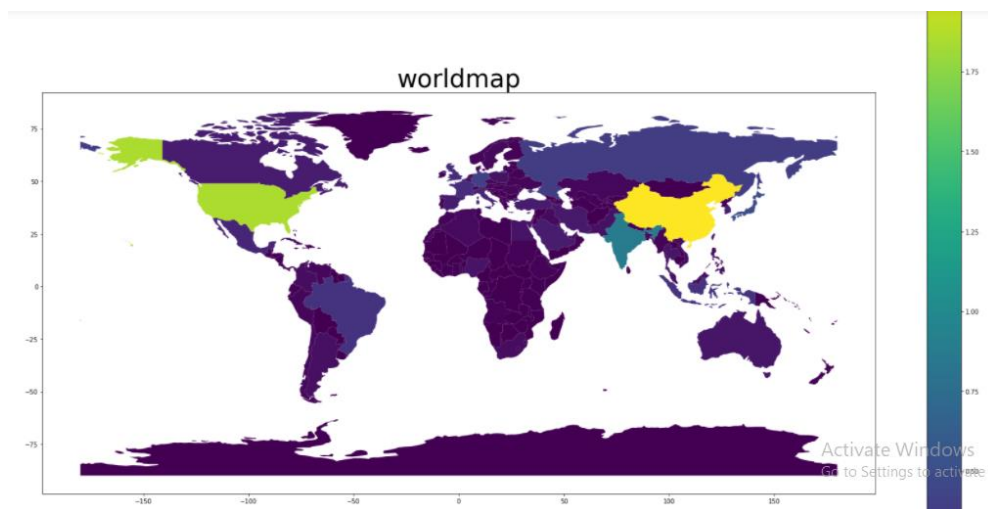
Dalam penelitian ini, kami menggunakan data yang merupakan data sekunder yang berupa data mengenai negara-negara. Terdapat 177 negara dengan 6 kolom data negara di seluruh dunia dan di sortir menjadi 47 data negara dan 6 kolom data negara di benua Asia. Clustering dengan pendekatan hirarki atau hierarchical clustering terlebih dahulu dilakukan pengelompokkan terhadap dua atau lebih objek yang memiliki karakteristik atau kemiripan paling dekat. Sedangkan clustering dengan pendekatan non hirarki atau partisi atau partitional clustering menentukan terlebih dahulu jumlah kelompok atau cluster yang diinginkan. Setelah itu, proses clustering dilakukan untuk mengetahui karakteristik (kemiripan) yang ada pada masing-masing cluster. Metode hirarki yang digunakan adalah Agglomerative Hierarchical Clustering (AHC) untuk mengklasifikasikan data yang sudah di label.

Terdapat beberapa penelitian terdahulu yang terkait dengan penelitian ini, antara lain: (1) menentukan sumber pasokan kopi secara optimal dengan menggunakan algoritma hierarchical clustering [5]; (2) implementasi algoritma hierarchical clustering untuk integrasi teks otomatis [6], [7]; (3) implementasi algoritma agglomerative hierarchical clustering untuk klasterisasi data kredit bank [8]; (4) pengelompokkan komunitas media sosial dengan algoritma agglomerative hierarchical clustering [9]; (5) penerapan algoritma hierarchical clustering untuk data runtun waktu [10]; dan (6) klasterisasi jenis cerita pendek pada media koran dengan menggunakan agglomerative hierarchical clustering [11].

## 2 Metodologi

Di bagian ini, kami mendeskripsikan mengenai pengaplikasian atau implementasi algoritma Agglomerative Clustering dalam mengklasterisasi atau memetakan negara-negara berdasarkan populasi dan GDP dari tiap negara di benua Asia.

Dataset yang digunakan dalam model latih ini diperoleh dari situs resmi [www.naturalearthdata.com](http://www.naturalearthdata.com). Data ini berisikan data dan kondisi dari tiap negara. Data ini dijadikan data latih untuk dilakukan klasterisasi menggunakan algoritma Agglomerative Clustering [12]–[14]. Model tersebut kemudian ditampilkan menggunakan library geopandas sehingga tampilan akan terlihat pengelompokan dari tiap-tiap negara. Klasterisasi ini dibatasi hanya data yang berada di benua Asia.



Gambar 1. Representasi skema dari model pelatihan

Dalam proses nya, data ini diolah dengan beberapa tahap, pertama dilakukan tahap processing dengan melakukan sortir terhadap data-data yang diperlukan. Kedua dilakukan pengkodean untuk Dilakukan klasterisasi dan terakhir visialisasi yang dapat menampilkan hasil klasterisasi.

## 3 Hasil dan Pembahasan

Dalam penelitian ini, kami menggunakan data yang merupakan data sekunder yang berupa data mengenai negara-negara. Terdapat .177 negara dengan 6 kolom data negara di seluruh dunia dan di sortir menjadi 47 data negara dan 6 kolom data negara di benua asia. Data tersebut terdiri dari id, nama, benua, populasi, GDP, dan letak geografis. Data tersebut ditunjukkan pada Gambar 2.

	iso_a3	name	continent	pop_est	gdp_md_est	geometry
0	AFG	Afghanistan	Asia	34124811.0	64080.0	POLYGON ((61.21082 35.65007, 62.23065 35.27066...
1	AGO	Angola	Africa	29310273.0	189000.0	MULTIPOLYGON (((23.90415 -11.72228, 24.07991 -...
2	ALB	Albania	Europe	3047987.0	33900.0	POLYGON ((21.02004 40.84273, 20.99999 40.58000...
3	ARE	United Arab Emirates	Asia	6072475.0	667200.0	POLYGON ((51.57952 24.24550, 51.75744 24.29407...
4	ARG	Argentina	South America	44293293.0	879400.0	MULTIPOLYGON (((-66.95992 -54.89681, -67.56244...
...	...	...	...	...	...	...
172	VUT	Vanuatu	Oceania	282814.0	723.0	MULTIPOLYGON (((167.84488 -16.46633, 167.51518...
173	YEM	Yemen	Asia	28036829.0	73450.0	POLYGON ((52.00001 19.00000, 52.78218 17.34974...
174	ZAF	South Africa	Africa	54841552.0	739100.0	POLYGON ((19.89577 -24.76779, 20.16573 -24.917...
175	ZMB	Zambia	Africa	15972000.0	65170.0	POLYGON ((23.21505 -17.52312, 22.56248 -16.898...
176	ZWE	Zimbabwe	Africa	13805084.0	28330.0	POLYGON ((29.43219 -22.09131, 28.79466 -21.639...

177 rows x 6 columns

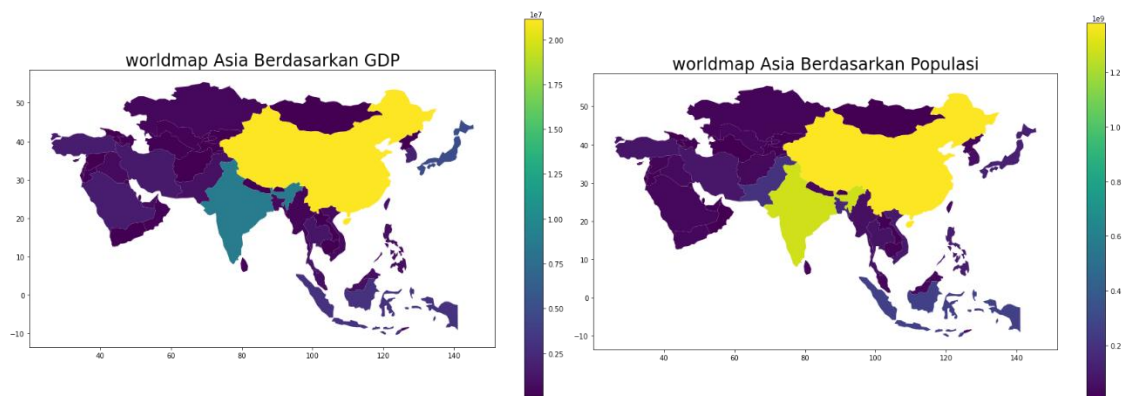
Gambar 2. Kumpulan Data di seluruh negara

	iso_a3	name	continent	pop_est	gdp_md_est	geometry
0	AFG	Afghanistan	Asia	34124811.0	64080.0	POLYGON ((61.21082 35.65007, 62.23065 35.27066...
3	ARE	United Arab Emirates	Asia	6072475.0	667200.0	POLYGON ((51.57952 24.24550, 51.75744 24.29407...
5	ARM	Armenia	Asia	3045191.0	26300.0	POLYGON ((43.58275 41.09214, 44.97248 41.24813...
10	AZE	Azerbaijan	Asia	9961396.0	167900.0	MULTIPOLYGON (((46.50572 38.77061, 46.48350 39...
15	BGD	Bangladesh	Asia	157826578.0	628400.0	POLYGON ((92.67272 22.04124, 92.65226 21.32405...
...	...	...	...	...	...	...
162	TUR	Turkey	Asia	80845215.0	1670000.0	MULTIPOLYGON (((43.58275 41.09214, 43.75266 40...
163	TWN	Taiwan	Asia	23508428.0	1127000.0	POLYGON ((121.77782 24.39427, 121.17563 22.790...
169	UZB	Uzbekistan	Asia	29748859.0	202300.0	POLYGON ((67.83000 37.14499, 67.07578 37.35614...
171	VNM	Vietnam	Asia	96160163.0	594900.0	POLYGON ((102.17044 22.46475, 102.70699 22.708...
173	YEM	Yemen	Asia	28036829.0	73450.0	POLYGON ((52.00001 19.00000, 52.78218 17.34974...

47 rows x 6 columns

Gambar 3. Kumpulan Data di Benua Asia

Pada tahap preprocessing ini kita dapat melihat daftar negara-negara di benua asia yang memiliki tingkat populasi dan GDP mulai dari terendah sampai tertinggi yang ditampilkan menggunakan worlmap menggunakan library geopandas worldmap classify. Tampilan data tersebut ditunjukkan pada Gambar 3.



Gambar 3. Proses Training

Dari tampilan data diatas, diketahui bahwa wilayah negara tertinggi berdasarkan GDP yaitu China dengan ditampilkan dengan wilayah warna kuning dengan jumlah 21140000, begitupula dengan nilai Populasi dengan jumlah 1379302771 orang. Kemudian negara yang memiliki nilai GDP dan populasi terkecil yaitu N Cyprus dengan nilai 3600 dan 256 ribu orang.

```
print('highest overall gdp in Asia:')
Asia.loc[Asia['gdp_md_est'].idxmax()]

highest overall gdp in Asia:
iso_a3          CHN
name            China
continent       Asia
pop_est        1.3793e+09
gdp_md_est      2.114e+07
geometry        POLYGON ((75.15802778514092 37.13303091078912...
Name: 30, dtype: object

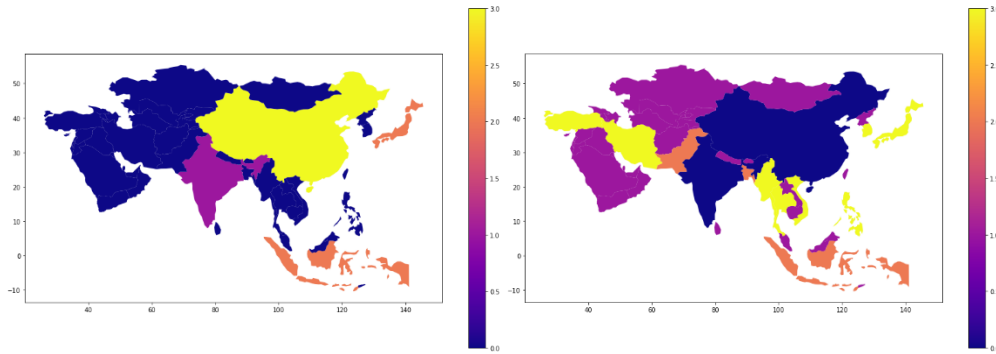
print('highest overall gdp in Asia:')
Asia.loc[Asia['gdp_md_est'].idxmin()]

highest overall gdp in Asia:
iso_a3          CYN
name            N. Cyprus
continent       Asia
pop_est        265100
gdp_md_est      3600
geometry        POLYGON ((32.73178022637745 35.14002594658844,...
Name: 38, dtype: object
```

Gambar 4. Training Loss dan Accuracy dataset

Kemudian pengkodean sebagai langkah implementasi algoritma agglomerative dalam mengklasifikasikan negara-negara di benua asia berdasarkan GDP dan Populasi sehingga menghasilkan nilai GDP perkapita pada negara. Pengkodean ini dilakukan menggunakan jupyter notebook dengan Bahasa python.

Klasterisasi pertama berdasarkan GDP dan Populasi negara di asia dengan masing-masing dibuat sebanyak 4 klaster.



Gambar 5. Pengujian

Pendapatan negara perkapita diperoleh dengan pembagian antara GDP dengan Populasi. Kemudian dibuat 4 klaster. Klaster 1 terdiri atas negara United Arab Emirates, Kuwait, dan Qatar. Klaster 2 terdiri atas Brunei, Oman, Saudi Arabia, dan Taiwan. Klaster 3 terdiri atas 34 negara dan klaster terakhir yaitu Cyprus, Israel, Kazakhstan, South Korea, Malaysia. Daftar negara yang termasuk dalam klaster 3 dapat dilihat pada Tabel 1. Sedangkan hasil dari seluruh data dapat dilihat pada Tabel 2.

Tabel 1. Daftar Negara di Asia

Afghanistan	India	Myanmar	Tajikistan
Armenia	Iran	Mongolia	Turkmenistan
Azerbaijan	Iraq	Nepal	Timor-Leste
Bangladesh	Jordan	Pakistan	Turkey
Bhutan	Kyrgyzstan	Philippines	Uzbekistan
China	Cambodia	North Korea	Vietnam
N. Cyprus	Laos	Palestine	Yemen
Georgia	Lebanon	Syria	
Indonesia	Sri Lanka	Thailand	

Tabel 2. Hasil Klasterisasi

iso_a3	name	pop_est	gdp_md_est	gdp/pop	cluster
AFG	Afghanistan	34.124.811	64080	0,001877813	2
ARE	United Arab Emirates	6.072.475	667200	0,109872828	0
ARM	Armenia	3.045.191	26300	0,008636568	2
AZE	Azerbaijan	9.961.396	167900	0,016855067	2
BGD	Bangladesh	157.826.578	628400	0,003981585	2
BRN	Brunei	443.593	33730	0,07603817	1
BTN	Bhutan	758.288	6432	0,008482265	2

iso_a3	name	pop_est	gdp_md_est	gdp/pop	cluster
CHN	China	1.379.302.771	21140000	0,015326584	2
CYN	N. Cyprus	265.100	3600	0,013579781	2
CYP	Cyprus	1.221.549	29260	0,023953194	3
GEO	Georgia	4.926.330	37270	0,00756547	2
IDN	Indonesia	260.580.739	3028000	0,011620199	2
IND	India	1.281.935.911	8721000	0,006802992	2
IRN	Iran	82.021.564	1459000	0,017788005	2
IRQ	Iraq	39.192.111	596700	0,015225003	2
ISR	Israel	8.299.706	297000	0,0357844	3
JOR	Jordan	10.248.069	86190	0,008410365	2
JPN	Japan	126.451.398	4932000	0,039003128	3
KAZ	Kazakhstan	18.556.698	460700	0,024826615	3
KGZ	Kyrgyzstan	5.789.122	21010	0,00362922	2
KHM	Cambodia	16.204.486	58940	0,003637264	2
KOR	South Korea	51.181.299	1929000	0,037689548	3
KWT	Kuwait	2.875.422	301100	0,104715064	0
LAO	Laos	7.126.706	40960	0,005747396	2
LBN	Lebanon	6.229.794	85160	0,013669794	2
LKA	Sri Lanka	22.409.381	236700	0,010562541	2
MMR	Myanmar	55.123.814	311100	0,005643659	2
MNG	Mongolia	3.068.243	37000	0,012059019	2
MYS	Malaysia	31.381.992	863000	0,027499848	3
NPL	Nepal	29.384.297	71520	0,002433953	2
OMN	Oman	3.424.386	173100	0,050549208	1
PAK	Pakistan	204.924.861	988200	0,004822255	2
PHL	Philippines	104.256.076	801900	0,007691638	2
PRK	North Korea	25.248.140	40000	0,001584275	2
PSX	Palestine	4.543.126	21220,77	0,004670962	2
QAT	Qatar	2.314.307	334500	0,144535708	0
SAU	Saudi Arabia	28.571.770	1731000	0,060584276	1
SYR	Syria	18.028.549	50280	0,00278891	2
THA	Thailand	68.414.135	1161000	0,016970177	2
TJK	Tajikistan	8.468.555	25810	0,003047745	2
TKM	Turkmenistan	5.351.277	94720	0,017700448	2
TLS	Timor-Leste	1.291.358	4975	0,003852534	2
TUR	Turkey	80.845.215	1670000	0,020656757	2
TWN	Taiwan	23.508.428	1127000	0,047940254	1
UZB	Uzbekistan	29.748.859	202300	0,006800261	2
VNM	Vietnam	96.160.163	594900	0,006186554	2
YEM	Yemen	28.036.829	73450	0,002619768	2

#### 4 Simpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan, maka dapat disimpulkan bahwa: Terdapat 4 kelompok atau cluster yang diperoleh dari analisis yang dilakukan. Adapun cluster tersebut adalah sebagai berikut: Klaster 1 terdiri atas negara United Arab Emirates, Kuwait, dan Qatar. Yang mana pada klaster ini dikategorikan negara dengan pendapatan perkapita tinggi dengan nilai GDP yang cukup besar. Klaster 2 terdiri atas Brunei, Oman, Saudi Arabia, dan Taiwan yang mana pada klaster

ini dikategorikan sebagai negara dengan pendapatan perkapita menengah ke tinggi. Klaster 3 terdiri atas 34 negara. Yang mana pada kluster ini dikategorikan sebagai negara dengan pendapatan per kapita bawah. Dan klaster terakhir yaitu Cyprus, Israel, Kazakhstan, South Korea, Malaysia dikategorikan sebagai negara dengan pendapatan per kapita menengah bawah.

## Referensi

- [1] Kompas, “Daftar Negara di Asia dan Ibu Kotanya,” *Kompas.com*, 2020. <https://www.kompas.com/skola/read/2020/01/06/120000369/daftar-negara-di-asia-dan-ibu-kotanya?page=all> (accessed Jan. 19, 2021).
- [2] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2006.
- [3] T. Alfina and B. Santosa, “Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Membentuk Cluster Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS),” *Anal. Perbandingan Metode Hierarchical Clust. K-means dan Gabungan Keduanya dalam Clust. Data*, vol. 1, no. 1, pp. 1–5, 2012.
- [4] R. Xu and D. C. Wunsch, *Clustering*. 2008.
- [5] A. Saputra, “OPTIMALISASI PENENTUAN SUMBER PASOKAN KOPI ARABIKA GAYO MELALUI PENDEKATAN HIRARCHICAL CLUSTERING DATA MINING,” *J. Optim.*, vol. 3, no. 4, 2018, doi: 10.35308/jopt.v3i4.214.
- [6] G. S. Budhi, A. I. Rahardjo, and H. Taufik, “Hierarchical Clustering untuk Aplikasi Automated Text Integration,” in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 2008, vol. 1, no. 1.
- [7] G. A. Pradnyana, “Perancangan dan Implementasi Automated Document Integration dengan Menggunakan Algoritma Complete Linkage Agglomerative Hierarchical Clustering,” *J. Ilmu Komput.*, vol. 5, no. 2, 2012.
- [8] G. Abdurrahman, “Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage,” *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 4, no. 1, p. 13, 2019, doi: 10.32528/justindo.v4i1.2418.
- [9] I. W. Rahayu, I. Atastina, and A. Herdiani, “Analisis Dan Implementasi Algoritma Agglomerative Hierarchical Clustering Untuk Deteksi Komunitas Pada Media Sosial Facebook,” *eProceedings Eng.*, vol. 5, no. 1, 2018.
- [10] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, “Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu,” *Jambura J. Math.*, vol. 1, no. 2, pp. 64–78, 2019, doi: 10.34312/jjom.v1i2.2354.
- [11] Z. Arifin, S. Santosa, and M. A. Soeleman, “Klasterisasi Genre Cerpen Kompas Menggunakan Agglomerative Hierarchical Clustering-Single Linkage,” *J. Cyberku*, vol. 13, no. 2, p. 2, 2017.
- [12] P. Contreras and F. Murtagh, “Hierarchical clustering,” in *Handbook of Cluster Analysis*, 2015, pp. 103–124.
- [13] L. Billard and E. Diday, “Agglomerative Hierarchical Clustering,” in *Clustering Methodology for Symbolic Data*, 2019, pp. 261–316.
- [14] R. Feldman and J. Sanger, “Hierarchical Agglomerative Clustering (HAC),” in *The Text Mining Handbook*, 2007, pp. 87–88.