



INSTITUT TEKNOLOGI DAN BISNIS STIKOM AMBON

**Program Studi Sistem Informasi
2024**



“Pengelompokan Harga Saham Bank BRI Menggunakan Metode Agglomerative Hierarchical Clustering”

Nama Mahasiswa:

**CITRA PRATAMA
200102002**

Dosen Pembimbing:

**M. Ikbal Siami, S.Kom.,M.Cs
NIPY :1218048601**

**Program Studi Sistem Informasi
Fakultas Ilmu Komputer
2024**





LATAR BELAKANG



Analisis terhadap harga saham menjadi sangat penting bagi investor untuk membuat keputusan investasi yang tepat. Mengingat volatilitas dan kompleksitas pergerakan harga saham diperlukan pendekatan analitis yang efektif untuk memahami dan mengelompokkan data harga saham berdasarkan karakteristik pergerakannya.

Salah satu pendekatan yang dapat digunakan adalah metode pengelompokan (clustering) menggunakan algoritma Agglomerative Hierarchical Clustering.



TUJUAN

Tujuan dari project ini adalah untuk mengelompokkan harga saham Bank BRI berdasarkan histori stock price menggunakan alogaritma Agglomerative Hierarchical Clustering dengan dikelompokkannya harga saham, diharapkan dapat memperoleh wawasan mengenai pola pergerakan harga saham yang serupa dalam periode tertentu dan membantu investor dalam membuat keputusan investasi yang lebih baik.



MANFAAT

01

Memberikan wawasan dan pengetahuan mengenai pengelompokan harga saham sehingga memperoleh gambaran yang jelas.

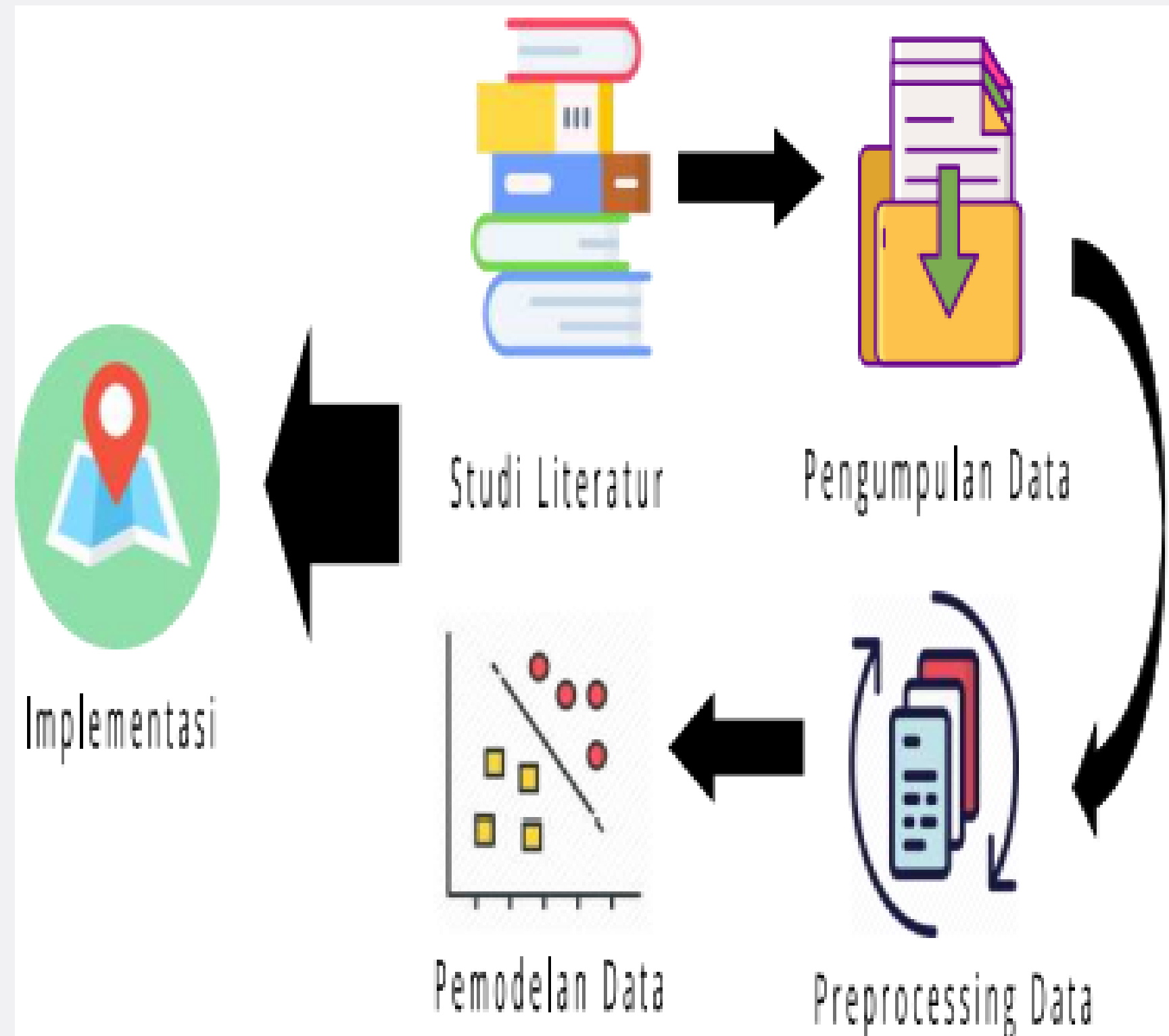
02

Pengambilan keputusan: Membantu investor dalam pengambilan keputusan berdasarkan kelompok harga saham dengan karakteristik yang mirip.

03

Strategi Investasi: Mengembangkan strategi informasi yang lebih baik berdasarkan pengelompokan harga saham

METODE PENELITIAN



PEMBAHASAN

CLUSTERING AGGLOMERATIVE

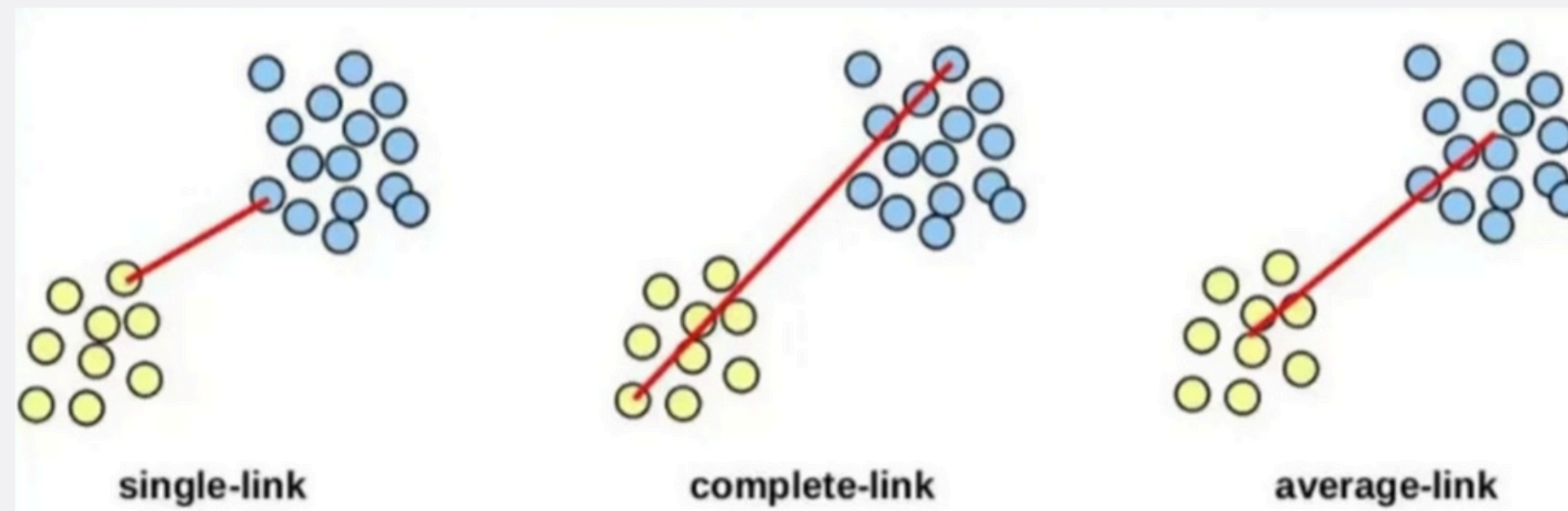
Strategi untuk pengelompokan hirarki bottom up, dimulai dari masing-masing data sebagai kluster kemudian membentuk kluster yang semakin besar.

METODE LINKAGE (Single Linkage, Average Linkage, Complete Linkage)

Single : Teknik penggabungan cluster jarak antara member terdekat dari dua cluster

Complete : Maximum distance between cluster

Average : Average distance between cluster



Simulasi Perhitungan Manual Euclidean Distance :

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

$$D(1,1), D(2,2), D(3,3), D(4,4), D(5,5)=0$$

$$D(1,2) = \sqrt{(1-4)^2 + (1-1)^2} = \sqrt{9} = 3$$

$$D(1,3) = \sqrt{(1-1)^2 + (1-2)^2} = \sqrt{1} = 1$$

$$D(1,4) = \sqrt{(1-3)^2 + (1-4)^2} = \sqrt{13} = 3,61$$

$$D(1,5) = \sqrt{(1-5)^2 + (1-4)^2} = \sqrt{25} = 5$$

$$D(2,3) = \sqrt{(4-1)^2 + (1-2)^2} = \sqrt{10} = 3,16$$

$$D(2,4) = \sqrt{(4-3)^2 + (1-4)^2} = \sqrt{10} = 3,16$$

$$D(2,5) = \sqrt{(4-5)^2 + (1-4)^2} = \sqrt{10} = 3,16$$

$$D(3,4) = \sqrt{(1-3)^2 + (2-4)^2} = \sqrt{8} = 2,82$$

$$D(3,5) = \sqrt{(1-5)^2 + (2-4)^2} = \sqrt{20} = 4,47$$

$$D(4,5) = \sqrt{(3-5)^2 + (4-4)^2} = \sqrt{4} = 2$$

Hierarchical Clustering (Complete Linkage)

Hasil Matrix Jarak

Dman	1	2	3	4	5
1	0	3	1	3,61	5
2	3	0	3,16	3,16	3,16
3	1	3,16	0	2,82	4,47
4	3,61	3,16	2,82	0	2
5	5	3,16	4,47	2	0

Data Set

Data	Fitur x	Fitur y
1	1	1
2	4	1
3	1	2
4	3	4
5	5	4

Pilih jarak kelompok terkecil 13 kemudian setelah dicari kelompok yang tersisa adalah jarak kelompok terbesar

Min (D_{man})=min(d₁₃)=1

D₍₁₃₎₂=max{d₁₂,d₃₂}=max{(3), (3,16)}=3,16

D₍₁₃₎₄=max{d₁₄,d₃₄}=max{(3,61), (2,28)}=3,61

D₍₁₃₎₅=max{d₁₅,d₃₅}=max{(5), (4,47)}=5

Dman	(13)	2	4	5
(13)	0	3,16	3,61	5
2	3,16	0	3,16	3,16
4	3,61	3,16	0	2
5	5	3,16	2	0

Pilih jarak kelompok terkecil $\text{Min}(D_{\text{man}}) = \min(d_{45}) = 2$

Hitung jarak kelompok (4,5) terhadap yang tersisa (13) dan 2

$$D_{(45)(13)} = \max\{d_{41}, d_{43}, d_{51}, d_{53}\} = \max\{(3,61), (2,82), (5), (4,47)\} = 5$$

$$D_{(45)2} = \max\{d_{42}, d_{52}\} = \max\{(3,16), (3,16)\} = 3,16$$

Dman	(45)	(13)	2
(45)	0	5	3,16
(13)	5	0	3,16
2	3,16	3,16	0

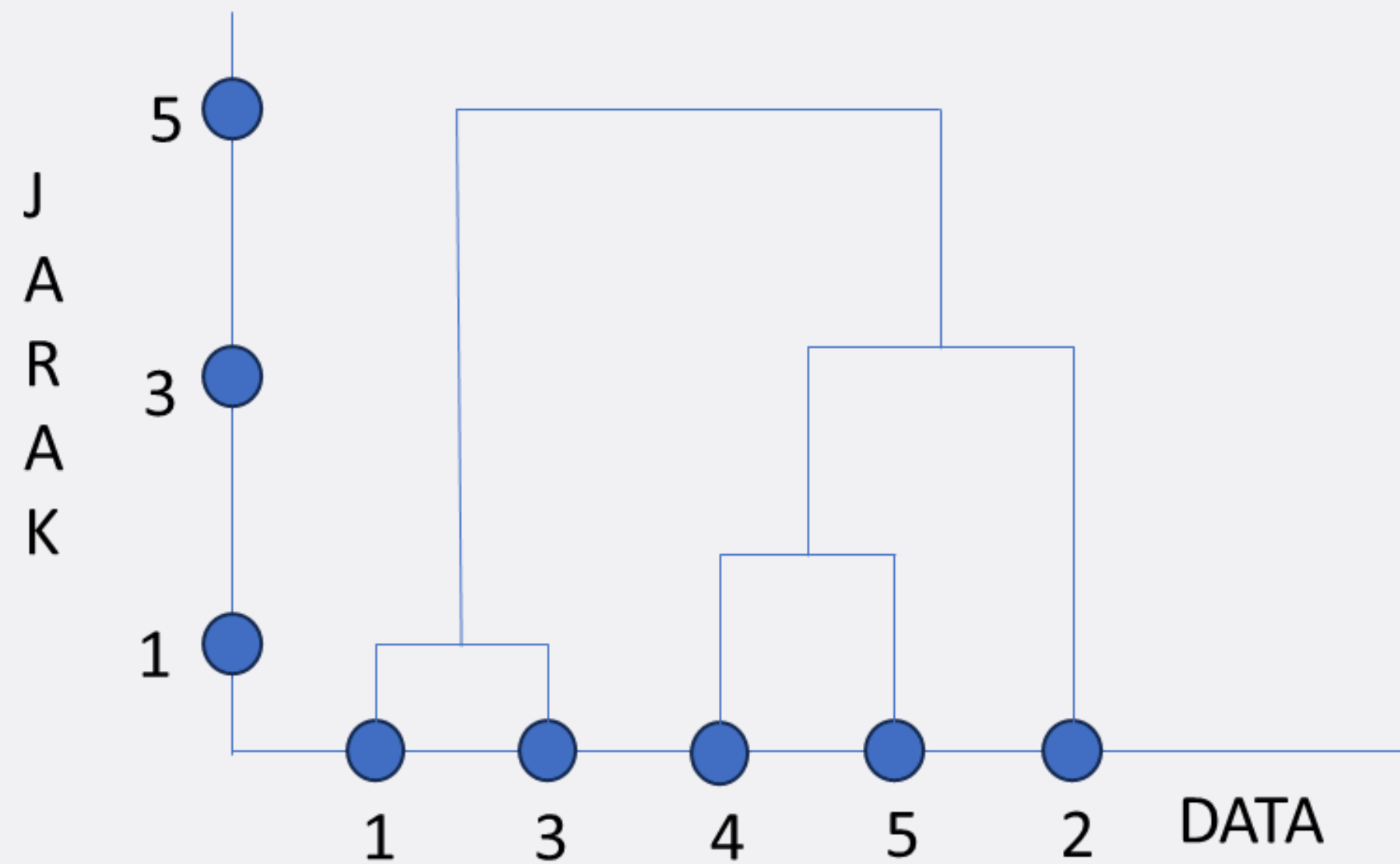
$$\text{Min}(D_{\text{man}}) = \min(d_{45}) = 3,16$$

Hitung jarak kelompok (45) dan 2 terhadap yang tersisa (13)

$$D_{(452)(13)} = \max\{d_{41}, d_{43}, d_{51}, d_{53}, d_{21}, d_{23}\} = \max\{(3,61), (2,82), (5), (4,47), (3), (3,16)\} = 5$$

Dman	(452)	(13)
(452)	0	5
(13)	5	0

Hasil Dendrogram Complete Linkage dengan jarak terjauh 5





Evaluasi Silhouette Score

Salah Satu cara yang dapat digunakan untuk mengetahui kekuatan cluster dan melihat kualitasnya adalah menggunakan Silhouette Score Rentang nilainya adalah -1 hingga +1.

Interpretasi Nilai Silhouette Score

Silhouette Score	Interpretasi
0.71 - 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak Terstruktur



Perangkat Lunak

Bahasa pemrograman python merupakan salah satu bahasa pemrograman yang menjadi pilihan untuk mengembangkan Machine Learning

Library Python

```
import pandas as pd
import numpy as np
import streamlit as st
import pickle
from sklearn.preprocessing import StandardScaler, Normalizer
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_score
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage
```



Fitur Streamlit

Streamlit adalah sebuah framework berbasis Python dan bersifat open-source yang dibuat untuk memudahkan dalam membangun aplikasi web di bidang sains data dan machine learning yang interaktif . Salah satu hal menarik dari framework ini adalah kita tidak perlu mengetahui banyak hal tentang teknologi web development. Kita tidak perlu dipusingkan tentang bagaimana mengatur tampilan website dengan CSS, HTML, atau Javascript. Untuk menggunakan Streamlit, kita cukup memiliki modal dasar mengetahui bahasa Python saja.



HASIL

Hierarchical Clustering (Complete Linkage)

Data Set

No	Date	Open	High	Low	Close	Volume	AdjustedClose
1	20240102	5650	5675	5625	5675	91134100	5675
2	20240103	5625	5650	5600	5600	83659700	5600
.....							
.....							
104	20240621	4280	4490	4260	4440	542704300	4440
105	20240624	4440	4510	4400	4400	303830800	4440



Hasil Clustering

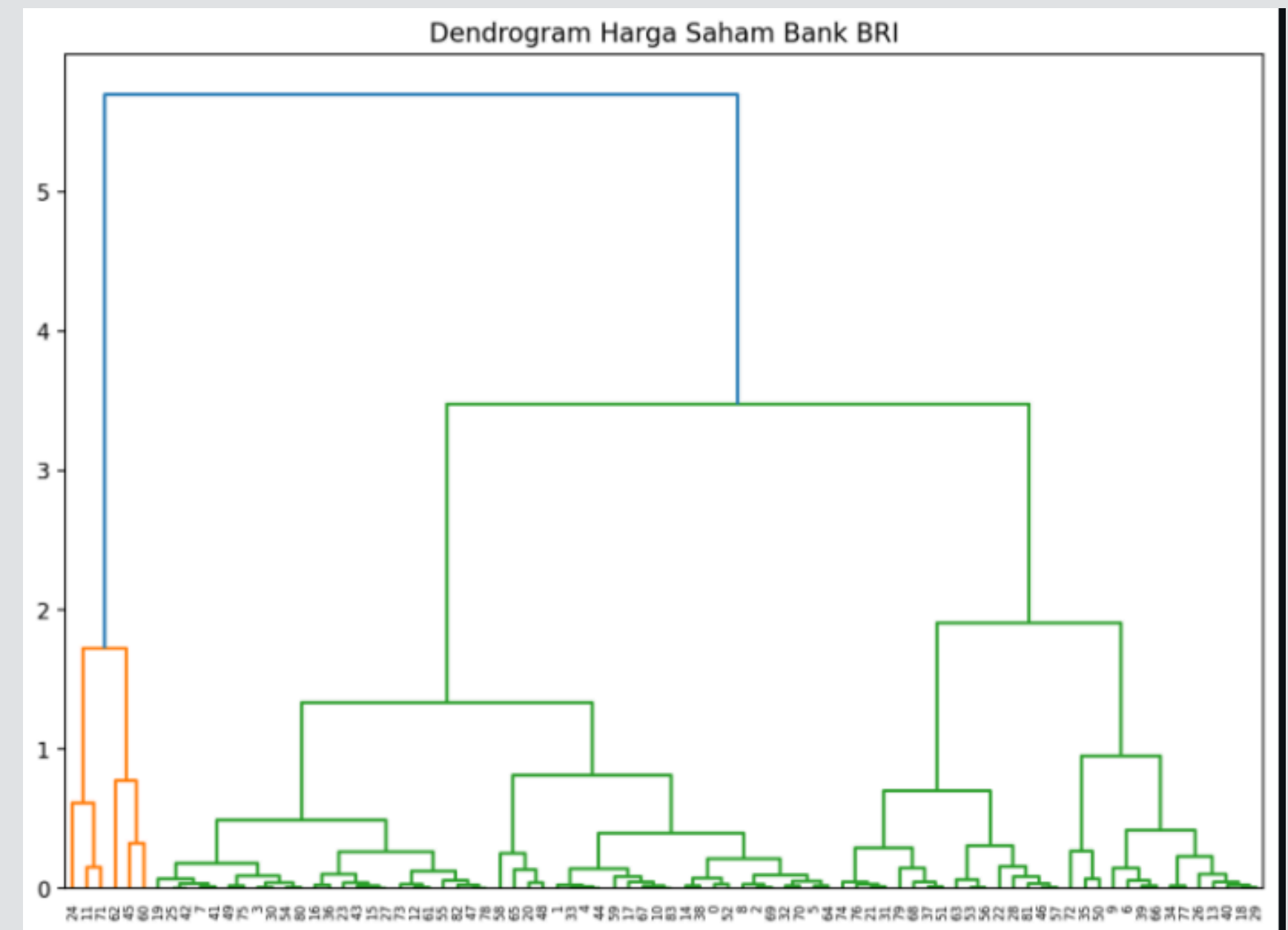
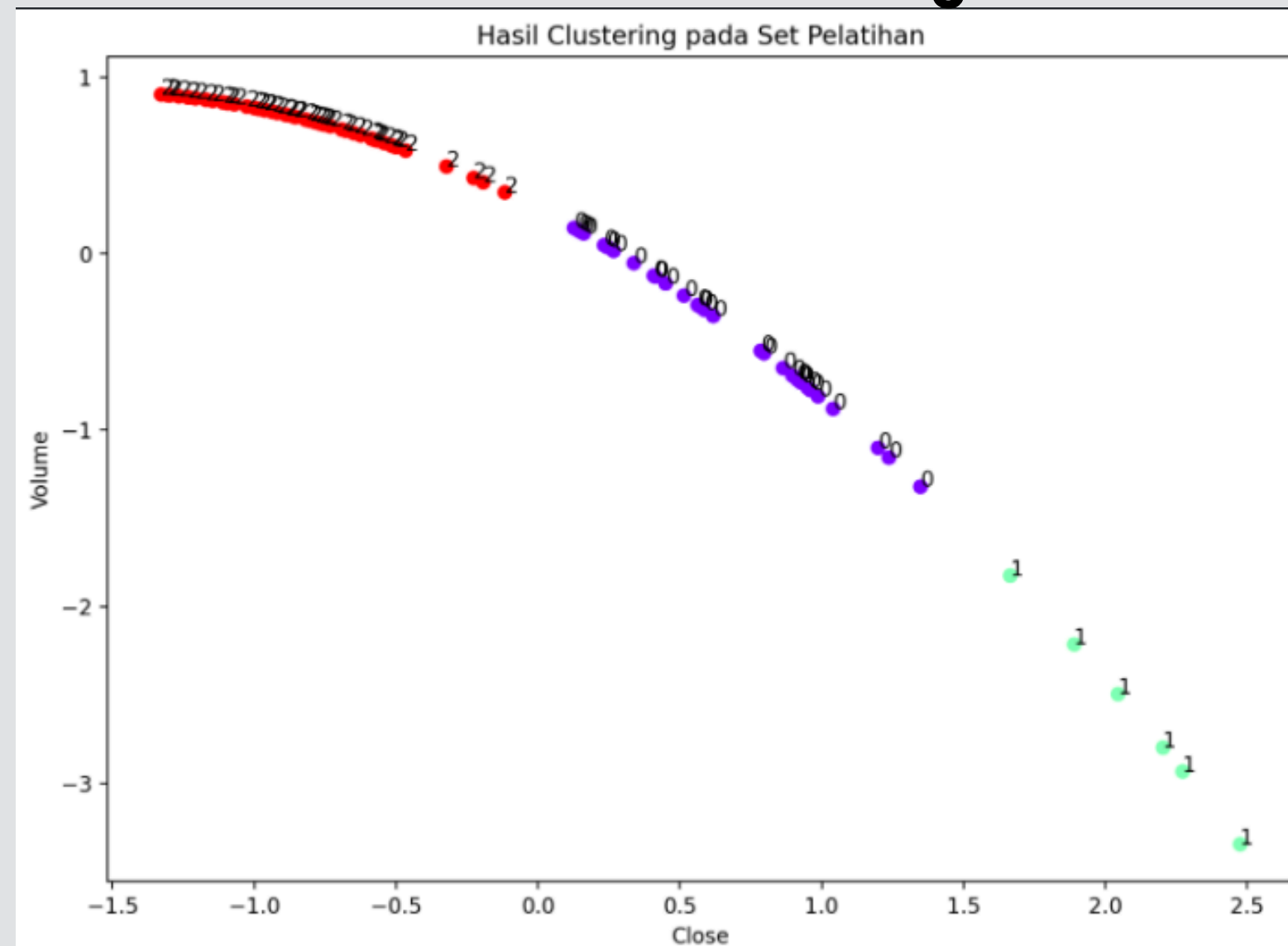
Cluster	Jumlah	Anggota Cluster
0	30	21,74,31,76,68,37,51,22,28,81,46,57,79,63,53,56,72,35,50,26,34,77,6,9,13,18,29,66,39,40
1	6	24,11,71,62,45,60
2	48	54,80,3,30,19,41,49,75,7,25,42,16,36,23,27,25,43,73,12,61,82,55,47,78,65,58,20,48,33,44,1,4,59,17,67,10,83,14,38,0,52,8,5,2,69,32,64,70

Data test sebanyak 20% dibentuk 3 cluster dengan menggunakan data sebanyak 84 data training, setiap satu cluster dicari jarak terpendek (yang paling mirip) dengan cluster lainnya, kemudian digabungkan menjadi satu. setelah itu cluster yang baru dihitung lagi jaraknya dengan cluster lain, dicari terjauh, kemudian digabungkan sampai akhirnya menjadi satu cluster.

Hasil Pembentukan Dendrogram

Hasil dari perhitungan jarak menggunakan perhitungan complete Euclidean dan Complete Linkage untuk mencari jarak kedekatan dari seluruh data yang dihitung, diperoleh pembentukan 3 cluster dalam membentuk dendrogram dengan Silhouette Score (Pelatihan): 0.70901627711322.

Scatter Plot Clustering




Silhouette Score (Pelatihan): 0.70901627711322



KESIMPULAN

Dari hasil pembahasan dan analisis yang telah dilakukan dapat ditarik kesimpulan:

1. Agglomerative Hierarchical Clustering dapat digunakan untuk mengelompokkan harga saham berdasarkan karakteristiknya dan dikaitkan dengan strategi investasi.
 2. Hasil pengelompokkan data sebanyak 85 data training diperoleh cluster 0 terdapat 30 anggota, cluster 1 dengan 6 anggota dan cluster 2 dengan 48 anggota.
 3. Evaluasi hasil metode Silhouette Score mendapatkan hasil yang jika dilihat pada tabel interpretasi nilai score dikatakan baik dan terstruktur yaitu (Pelatihan): 0.70901627711322.
 4. Hasil dari analisis ini dapat membantu investor dalam membuat keputusan investasi yang lebih baik dan terinformasi.
- 

SARAN

Berdasarkan hasil penelitian yang diperoleh beberapa saran yang dapat meningkatkan efektivitas proyek pengelompokan harga saham Bank Mandiri menggunakan algoritma Agglomerative Hierarchical Clustering:

- Streamlit Dashboard: Kembangkan dashboard interaktif menggunakan Streamlit yang memungkinkan pengguna untuk unggah data, melakukan clustering, dan melihat hasil serta interpretasi secara real-time.
- Integrasi API: Pertimbangkan integrasi dengan API untuk mendapatkan data saham real-time dan memperbarui hasil clustering secara berkala.



TERIMA KASIH

