

# Implementasi Graf pada Algoritma Klusterisasi (*Clustering*)

Muh. Muslim Al Mujahid – 13518054

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

13518054@std.stei.itb.ac.id

**Abstract**—Pengelompokan (*clustering*) adalah salah satu metode pemrosesan data dengan pengaplikasian yang sangat luas. Konsep ini banyak digunakan pada analisis data *market* dan juga diaplikasikan sebagai salah satu pendekatan *machine learning* (*unsupervised learning*).

**Keywords**—*Graph, Clustering, Data Analysis*

## I. PENDAHULUAN

Beberapa dekade yang lalu, manusia mulai menyadari betapa pentingnya data. Oleh karena itu setiap lembaga dan organisasi mulai mengumpulkan sebanyak mungkin data untuk digunakan. Akan tetapi data-data ini tersimpan dengan berbagai bentuk, sehingga secara umum harus dilakukan suatu proses terlebih dahulu untuk dapat melihat informasi yang terkandung pada data tersebut.

Analisis Data (*Data Analysis*) didefinisikan sebagai proses pembersihan, transformasi, dan pemodelan untuk mencari informasi yang berguna dari sebuah atau sekumpulan data. Analisis data banyak digunakan pada perusahaan bisnis untuk melakukan *decision-making*, karena pada umumnya data berukuran besar sehingga tidak mungkin dilakukan pengecekan secara manual ke tiap data individu. Salah satu metode analisis data yang umum digunakan adalah pengelompokan (*Clustering*).

*Clustering* adalah suatu tugas untuk mengelompokkan populasi atau poin-poin data ke beberapa kelompok sehingga poin data pada suatu kelompok lebih mirip satu sama lain dibandingkan dengan poin data pada kelompok lainnya.

Dengan melakukan *clustering* pada sebuah data, kita dapat melihat bahwa beberapa data dapat dibagi menjadi beberapa kelompok. Contohnya sebuah toko mengumpulkan data para pembelinya dengan informasi umur dan rata-rata pembelian per bulan. Dengan melakukan *clustering* pada data ini, pelanggan dapat saja terbagi menjadi 4 kelompok yaitu tua dan pembelian sedikit, tua dan pembelian banyak, muda dan pembelian sedikit, dan muda dan pembelian banyak. Ini hanyalah contoh, data dapat saja terbagi menjadi kelompok yang lebih kecil atau lebih besar.

*Clustering* juga banyak diterapkan pada pengembangan kecerdasan buatan di bidang *machine learning*. *Clustering* termasuk ke dalam algoritma *unsupervised machine learning*, dimana algoritma mencari pola (*pattern*) pada sebuah data dengan melakukan pengelompokan (*Clustering*).

## II. TEORI DASAR

### A. Graf

Graf adalah kumpulan simpul-simpul yang dihubungkan dengan sisi. Graf digunakan untuk merepresentasikan objek-objek diskrit dan hubungan antara objek-objek tersebut. Graf terdiri dari 2 komponen, yaitu simpul dan sisi. Sedemikian sehingga jika diberikan graf  $G$  maka,

$$G = (V, E)$$

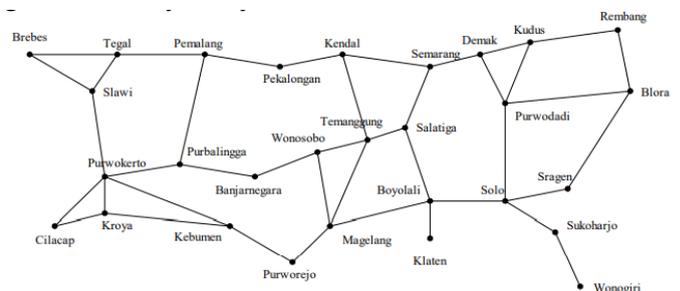
$V$  = Himpunan tidak kosong dari simpul

$$= \{v_1, v_2, v_3, \dots, v_n\}$$

$E$  = Himpunan tidak kosong dari sisi

$$= \{e_1, e_2, e_3, \dots, e_n\}$$

Gambar berikut merupakan contoh graf yang menyatakan peta jaringan jalan raya yang menghubungkan sejumlah kota di Provinsi Jawa Tengah.



Gambar 2.1: Contoh penerapan graf pada peta jaringan jalan raya yang menghubungkan sejumlah kota di Provinsi Jawa Tengah (Sumber: Graf-2020-Bagian1.pdf)

Setiap kota direpresentasikan sebagai simpul, dan jalan raya yang menghubungkan kota A dan kota B sebagai sisi.

### B. Jenis-Jenis Graf

Berdasarkan ada atau tidaknya gelang atau sisi ganda pada suatu graf, maka graf digolongkan menjadi 2 jenis

- Graf Sederhana (*Simple Graph*)  
Graf sederhana adalah graf yang tidak memiliki gelang maupun sisi ganda.

- Graf tak-sederhana (*Unsimple Graph*)  
Graf tak-sederhana adalah graf yang memiliki gelang atau sisi ganda, atau keduanya.

Berdasarkan ada atau tidaknya orientasi arah pada suatu graf, maka graf digolongkan menjadi 2 jenis

- Graf tak-berarah (*Undirected Graph*)  
Graf tak-berarah adalah graf yang tidak memiliki orientasi arah pada sisinya.
- Graf berarah (*Directed Graph*)  
Graf berarah adalah graf yang memiliki orientasi arah pada sisinya.

Berdasarkan ada atau tidaknya bobot pada suatu graf, maka graf digolongkan menjadi 2 jenis

- Graf tak-berbobot (*Unweighted Graph*)  
Graf tak-berbobot adalah graf yang tidak memiliki bobot pada simpulnya, atau untuk setiap sisi yang menghubungkan setiap simpul memiliki bobot yang sama (sehingga bisa diabaikan).
- Graf berbobot (*Weighted Graph*)  
Graf berbobot adalah graf yang memiliki bobot pada sisinya. Bobot disini dapat berupa jarak atau biaya yang diperlukan untuk berpindah dari satu simpul ke simpul lainnya.

### C. Terminologi Pada Teori Graf

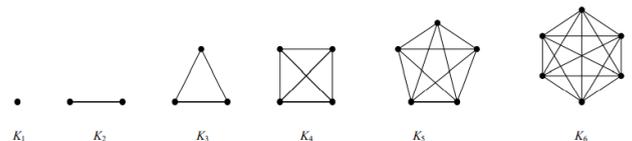
- Bertetangga (*Adjacent*)  
Dua buah simpul  $v_j$  dan  $v_k$  dikatakan bertetangga jika terdapat satu atau lebih sisi sehingga  $e_i = (v_j, v_k)$  menghubungkan  $v_j$  dan  $v_k$ .
- Bersisian (*Incidency*)  
Untuk sembarang sisi  $e = (v_j, v_k)$ ,  $e$  dikatakan bersisian dengan simpul  $v_j$  dan  $v_k$  karena menghubungkan simpul  $v_j$  dan  $v_k$  atau sebaliknya.
- Simpul terpencil (*Isolated Vertex*)  
Simpul terpencil adalah simpul yang tidak memiliki sisi yang bersisian dengannya
- Graf kosong (*null graph* atau *empty graph*)  
Graf kosong adalah graf yang tidak memiliki sisi atau himpunan sisinya ( $V$ ) adalah himpunan kosong.
- Derajat (*Degree*)  
Derajat sebuah simpul  $v$  adalah  $d(v)$ , yaitu jumlah sisi yang bersisian dengan simpul  $v$ .
- Lintasan (*Path*)  
Lintasan adalah barisan berselang-seling simpul-simpul

dan sisi yang berbentuk  $v_0, e_1, v_1, e_2, v_2, \dots, v_{n-1}, e_n, v_n$  sedemikian sehingga  $e_1 = (v_0, v_1), e_2 = (v_1, v_2), \dots, e_n = (v_{n-1}, v_n)$  adalah sisi-sisi pada graf  $G$ , yang secara berurutan menghubungkan simpul-simpul yang saling bertetangga. Panjang sebuah lintasan adalah jumlah sisi yang dilalui pada lintasan tersebut untuk graf tak-berbobot, atau jumlah bobot dari setiap sisi yang dilewati pada lintasan untuk graf berbobot.

- Siklus (*Cycle*) atau Sirkuit (*Circuit*)  
Siklus atau sirkuit adalah lintasan yang berawal dan berakhir pada simpul yang sama.
- Terhubung (*Connected*)  
Simpul  $v_1$  dan  $v_2$  dikatakan terhubung jika terdapat lintasan yang menghubungkan  $v_1$  dan  $v_2$  atau sebaliknya.

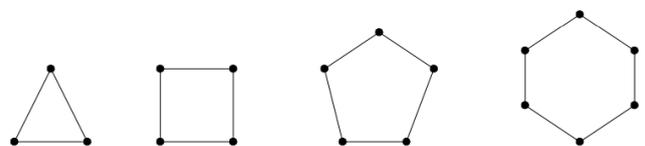
### D. Beberapa Graf Khusus

- Graf lengkap (*Complete graph*)  
Graf lengkap adalah graf sederhana yang untuk setiap simpulnya memiliki sisi yang menghubungkannya ke setiap simpul lainnya. Graf lengkap dengan  $n$  buah simpul dilambangkan dengan  $K_n$ . Jumlah sisi pada graf lengkap yang terdiri dari  $n$  buah simpul adalah  $n(n-1)/2$  sisi.



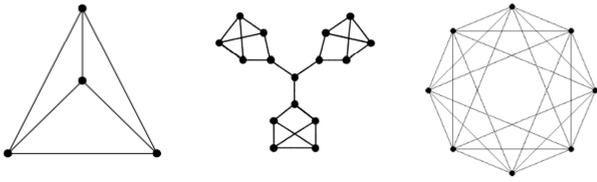
Gambar 2.2: Contoh graf lengkap (Sumber: Graf-2020-Bagian I.pdf)

- Graf lingkaran  
Graf lingkaran adalah graf sederhana yang setiap simpulnya berderajat dua. Graf lingkaran dengan  $n$  simpul dilambangkan dengan  $C_n$ .

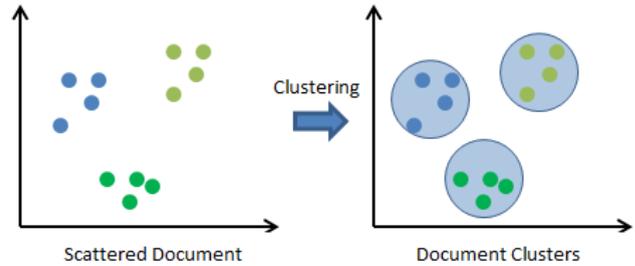


Gambar 2.3: Contoh graf lingkaran (Sumber: Graf-2020-Bagian I.pdf)

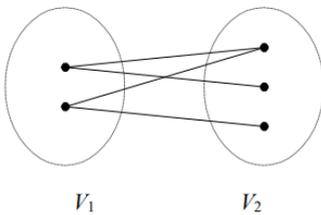
- Graf teratur  
Graf teratur adalah graf yang setiap simpulnya memiliki derajat yang sama.



Gambar 2.4: Contoh graf teratur (Sumber: Graf-2020-Bagian1.pdf)

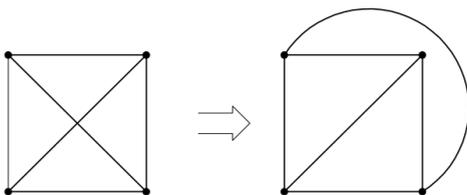


- Graf bipartite (*Bipartite graph*)  
 Graf bipartite adalah graf G yang himpunan simpulnya dapat dibagi menjadi dua himpunan bagian  $V_1$  dan  $V_2$ , sedemikian sehingga setiap sisi pada G menghubungkan sebuah simpul di  $V_1$  ke sebuah simpul  $V_2$  sehingga memenuhi syarat himpunan.



Gambar 2.5: Contoh graf bipartite (Sumber: Graf-2020-Bagian1.pdf)

- Graf planar dan graf bidang  
 Graf planar adalah graf yang dapat digambarkan pada bidang datar sedemikian sehingga tidak ada sisi yang saling berpotongan. Graf bidang adalah graf planar yang digambarkan sehingga tidak ada sisi yang saling berpotongan.



Gambar 2.5: Contoh graf planar dan graf bidang (Sumber: Graf-2020-Bagian1.pdf)

### E. Pengelompokan (*Clustering*)

Pengelompokan (*Clustering*) adalah suatu tugas untuk mengelompokkan populasi atau poin-poin data ke beberapa kelompok sehingga poin data pada suatu kelompok lebih mirip satu sama lain dibandingkan dengan poin data pada kelompok lainnya. Misal diberikan himpunan data P. Akan dicari himpunan data  $P_1$  dan  $P_2$  dimana  $P_1 \cup P_2 = P$  sedemikian sehingga  $a_i$  adalah elemen dari  $P_1$  dan  $b_i$  adalah elemen dari  $P_2$  sedemikian sehingga  $\|a_i - a_j\| < \|a_i - b_j\|$  dan sebaliknya.

## III. PENGELOMPOKAN (*CLUSTERING*) DENGAN MENGGUNAKAN GRAF

### A. Merepresentasikan Data Sebagai Graf

Untuk merepresentasikan sebuah data ke dalam graf, pertama harus ditentukan apa yang akan menjadi simpul dan sisi pada graf tersebut. Dalam kasus *clustering* ini, tiap data akan menjadi simpul pada graf, dan sisi menandakan keterhubungan antara satu simpul dengan simpul lain. Pada kasus *clustering* ini akan digunakan graf berbobot (*Weighted graph*). Bobot dari simpul  $v_i$  ke simpul  $v_j$  adalah jarak *euclidean* dari data ke-i dan data ke-j. Bobot sisi pada graf dapat dihitung dengan formula jarak *euclidean* sebagai berikut.

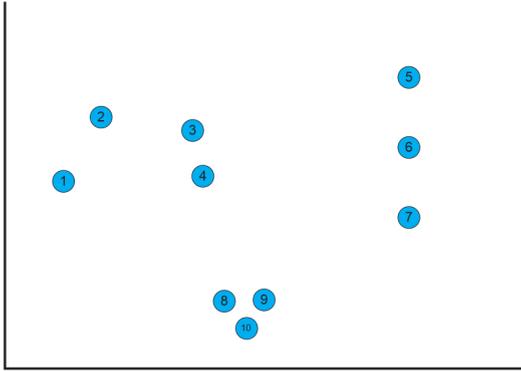
$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

dimana  $x_i$  adalah nilai poin data pertama untuk fitur ke-i dan  $y_i$  adalah poin data kedua untuk fitur ke-i. Semakin kecil bobot dari sebuah sisi menunjukkan bahwa keterhubungan data yang direpresentasikan sebagai simpul semakin kuat dan lebih besar kemungkinan akan berada pada klaster yang sama.

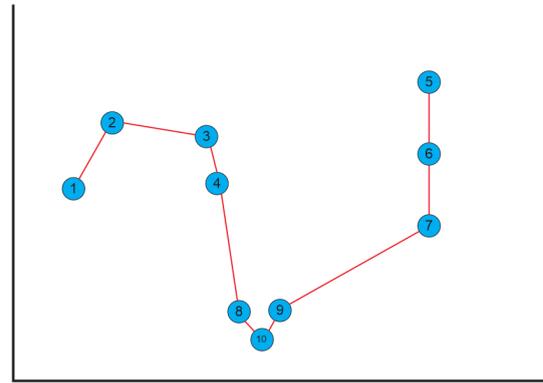
### B. Membangun Graf Dengan Data

Terdapat beberapa tahap yang harus dilakukan untuk membangun graf menggunakan data sebagai berikut.

- Pada awalnya, graf G merupakan graf kosong (*null graph* atau *empty graph*), yaitu graf yang tidak memiliki sisi, dengan simpul sebagai poin data atau graf kosong yang hanya memiliki satu simpul.



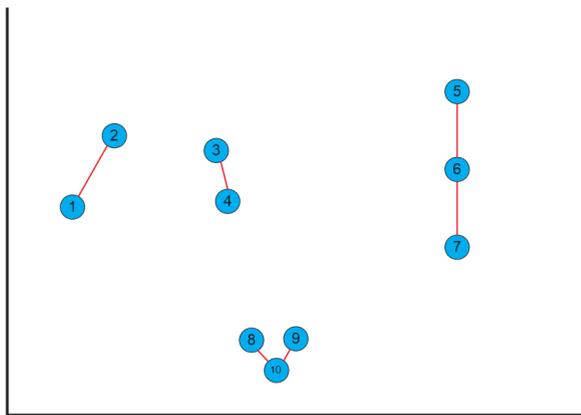
Gambar 3.1: Membangun graf dengan data iterasi ke-1



Gambar

3.3: Membangun graf dengan data iterasi ke-3

- Pada iterasi pertama, hubungkan setiap simpul ke simpul lainnya yang jaraknya paling dekat dengan simpul tersebut. Sehingga pada data diatas akan diperoleh  $e_1 = (v_1, v_2)$ ,  $e_2 = (v_3, v_4)$ ,  $e_3 = (v_5, v_6)$ ,  $e_4 = (v_6, v_7)$ ,  $e_5 = (v_8, v_{10})$ , dan  $e_6 = (v_9, v_{10})$  dengan visualisasi graf sebagai berikut.



Gambar 3.2: Membangun graf dengan data iterasi ke-2

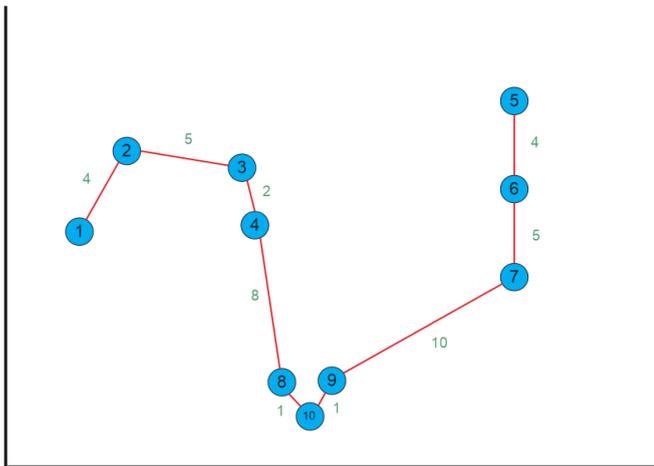
- Selanjutnya untuk setiap graf tak-terhubung baru yang terbentuk, hubungkan dengan graf lainnya melalui simpul dari kedua graf yang berjarak minimal. Sehingga diperoleh sisi baru  $e_7 = (v_2, v_3)$ ,  $e_8 = (v_4, v_8)$ , dan  $e_9 = (v_7, v_9)$ . Tahap ini terus dilakukan secara berulang-ulang sampai graf G menjadi graf terhubung.

### C. Pengelompokan (Clustering) Data Pada Graf

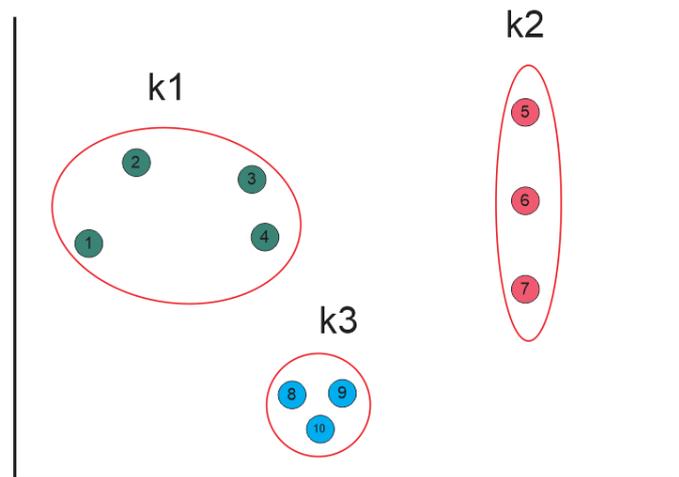
Misal untuk setiap sisi pada graf yang telah dibangun memiliki bobot sebagai berikut.

Sisi	Simpul yang terhubung	Bobot
$e_1$	$(v_1, v_2)$	4
$e_2$	$(v_3, v_4)$	2
$e_3$	$(v_5, v_6)$	4
$e_4$	$(v_6, v_7)$	5
$e_5$	$(v_8, v_{10})$	1
$e_6$	$(v_9, v_{10})$	1
$e_7$	$(v_2, v_3)$	5
$e_8$	$(v_4, v_8)$	8
$e_9$	$(v_7, v_9)$	10

Sehingga jika divisualisasikan akan menjadi sebagai berikut.



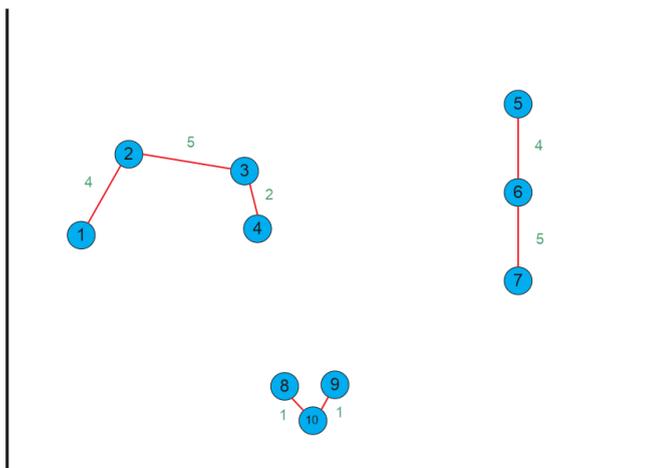
Gambar 3.4: Visualisasi graf data dengan bobot



Gambar 3.6: Hasil pewarnaan tiap kluster

Sebelum melakukan pengelompokan (*Clustering*), pertama kita harus menentukan nilai  $k$ , yaitu jumlah kluster yang ingin dihasilkan dari data yang diberikan. Misal kita ambil  $k = 3$ , maka graf akan dikelompokkan menjadi 3 kluster.

Untuk mendapatkan  $k$  kluster dari graf data, maka akan dilakukan penghapusan/menghilangkan sisi dengan bobot terbesar sebanyak  $k - 1$  kali. Sehingga jika  $k = 3$ , maka akan dilakukan penghapusan sisi dengan bobot terbesar sebanyak  $3 - 1 = 2$  kali. Pada graf data yang kita miliki, dua sisi yang memiliki bobot terbesar adalah  $e_8$  dan  $e_9$  dengan bobot 8 dan 10. Sehingga akan didapatkan graf-graf baru sebagai berikut.



Gambar 3.5: Hasil pengelompokan (*clustering*) pada graf data

Selanjutnya kita tinggal melakukan pewarnaan pada setiap graf yang dihasilkan. Setiap graf melambangkan satu kluster, dan simpul pada graf tersebut adalah anggota dari kluster tersebut.

Maka diperoleh ketiga kluster adalah  $C_1 = \{ v_1, v_2, v_3, v_4 \}$ ,  $C_2 = \{ v_5, v_6, v_7 \}$ , dan  $C_3 = \{ v_8, v_9, v_{10} \}$ .

#### IV. ANALISIS HASIL

Untuk algoritma yang sederhana bisa dibilang hasil yang diberikan sudah sangat bagus. Kekurangan dari algoritma ini adalah semakin banyak data maka graf yang dibentuk akan semakin besar yang tentu saja akan memakan banyak komputasi dan memori. Selain itu diperlukan sebuah metode khusus untuk mencari jumlah kluster paling optimal sehingga hasil pengelompokan menghasilkan informasi yang maksimal.

#### V. KESIMPULAN

Graf adalah salah satu konsep yang sangat banyak penerapannya di kehidupan kita sehari-hari, salah satunya adalah pengelompokan (*clustering*). Data dapat direpresentasikan sebagai graf dimana simpul merepresentasikan data poin secara individu dan sisi merepresentasikan keterhubungan antara satu data poin dengan data poin yang lain. Graf data berupa graf tak-berarah dan berbobot, dimana bobotnya dapat dihitung dengan menggunakan formula jarak *euclidean*. Semakin kecil bobot suatu sisi, maka semakin mirip pula data poin yang dihubungkan sehingga semakin besar kemungkinan berada pada kluster yang sama. Graf dibangun dengan melakukan iterasi menghubungkan tiap graf tak-terhubung dengan graf tak-berarah lainnya sampai menjadi satu graf terhubung. Pengelompokan dilakukan dengan memotong sisi dengan bobot paling besar. Setiap graf yang dihasilkan merupakan satu kluster dan data pada simpul termasuk ke dalam kluster tersebut.

## VI. UCAPAN TERIMA KASIH

Puji dan syukur saya panjatkan kehadirat Allah SWT. Karena rahmat dan karunia-Nya lah saya masih sempat untuk menyelesaikan makalah ini dengan tepat waktu. Terima kasih saya ucapkan kepada Bu Harlili dan seluruh dosen mata kuliah IF2120 Matematika Diskrit yang selama ini telah mengajar dan membimbing saya dalam penyelesaian makalah ini. Saya juga berterima kasih kepada kedua Orang tua saya yang atas doa dan dukungan yang selalu diberikan serta teman-teman yang telah membantu dalam penyelesaian makalah ini.

## REFERENCES

- [1] Munir, Rinaldi. Matematika Diskrit. 2010. Informatika Bandung: Bandung, Edisi 3
- [2] Kaushik, Sauravkaushik8. An Introduction to Clustering and different methods of clustering. 2016. <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/> (Diakses pada tanggal 14 Desember 2021 pukul 18.14 WIB)
- [3] Johnson, Daniel. What is Data Analysis? Research | Types | Methods | Techniques. 2021. <https://www.guru99.com/what-is-data-analysis.html> (Diakses pada tanggal 14 Desember 2021 pukul 12.47 WIB)

## PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Polewali Mandar, 14 Desember 2021



Muh. Muslim Al Mujahid  
13518054