

**FINAL PROJECT**  
**ANALISIS MULTIVARIAT (A)**



**Vanny Khairunnisaa**

**2206051506**

**Dosen Pengampu**

**Prof. Dr. Dra. Titin Siswantining, D.E.A.**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS INDONESIA**

**2024**

## DAFTAR ISI

<b>DAFTAR ISI</b>	<b>1</b>
<b>BAB I PENDAHULUAN</b>	<b>2</b>
<b>BAB II PRE-PROCESSING DAN ANALISIS DESKRIPTIF</b>	<b>4</b>
<b>BAB III METODE</b>	<b>7</b>
1. Menghitung matriks jarak	8
2. Membuat matriks A	8
3. Membuat matriks B	8
4. Mencari nilai eigen value dan eigen vector	9
5. Menghitung jarak koordinat yang terbentuk	9
6. Menghitung nilai STRESS	9
<b>BAB IV PENGOLAHAN DATA ANALISIS DAN HASIL</b>	<b>10</b>
1. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Jenis Hutan dan Cuaca	12
2. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Jenis Hutan	13
3. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Cuaca	14
4. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Tanah Hidup	15
5. Nilai STRESS	16
<b>BAB V PENUTUP</b>	<b>17</b>
<b>DAFTAR PUSTAKA</b>	<b>19</b>
<b>LAMPIRAN</b>	<b>20</b>

## BAB I PENDAHULUAN

Tanaman merupakan aspek penting dalam menjaga kestabilan ekosistem dan lingkungan alam. Sebagai makhluk hidup, tanaman memiliki berbagai sel yang berkaitan dengan unsur-unsur kimia. Konsentrasi bahan kimia dalam tanaman dapat dipengaruhi oleh beberapa faktor lingkungan seperti cuaca, hutan tempat tinggal, dan jenis tanah yang ada. Dengan memahami hubungan konsentrasi bahan kimia pada tanaman yang bervariasi antar spesies tanaman dengan berbagai faktor lingkungan seperti cuaca, hutan tempat tinggal, dan jenis tanah tanaman hidup dapat membantu proses identifikasi jenis tanaman berdasarkan tipe cuaca, tipe hutan, dan tipe jenis tanah tanaman pada saat itu.

Untuk mengidentifikasi konsentrasi bahan kimia pada tanaman berdasarkan faktor lingkungan diperlukan sebuah metode statistik. Terdapat berbagai teknik analisis statistik yang dapat digunakan, salah satu teknik analisis statistik multivariat yang dapat digunakan adalah *Multidimensional Scaling* (MDS). *Multidimensional Scaling* (MDS) digunakan untuk menggambarkan struktur hubungan antar objek data berdasarkan kemiripannya secara grafik dalam suatu bidang multidimensi untuk mendapatkan informasi data. Pengaplikasian *Multidimensional Scaling* (MDS) pada penelitian ini bertujuan untuk mengeksplorasi kemiripan dan perbedaan konsentrasi bahan kimia pada spesies tanaman di Oslo, Norwegia dengan berbagai faktor seperti kondisi cuaca, tipe hutan, dan jenis tanah tanaman.

Data yang digunakan pada penelitian ini berasal dari suatu *library* R yaitu “heplots” dengan data bernama “Oslo”. Data ini berisikan 332 observasi berupa berbagai jenis spesies tanaman yang diambil pada jarak radius 120 km mengelilingi kota Oslo, Norwegia. Berikut adalah cuplikan dari data

rownames	site	XC	YC	forest	weather	litho	altitude	Cu	Fe	K	Mg	Mn	P	Zn
2	102	582865	6706649	sprbir	nice	gneis_r	431	1.893111963	3.912023005	9.314700387	8.202482447	7.419979924	7.820037989	5.693732139
3	103	581155	6720326	sprbir	rain	gneis_r	430	1.637053079	3.912023005	9.179881164	8.045588281	7.857093865	7.820037989	5.257495372
4	104	580482	6711512	sprbir	moist	gneis_r	475	1.591273942	3.688879454	8.895629627	8.328451067	7.80139132	7.889833751	5.614222398
5	105	584618	6701621	spruce	moist	gneis_r	410	1.819698838	4.248495242	9.200290036	8.556413905	6.64509097	7.620705087	4.599152114
6	106	581453	6698513	birspr	nice	camsed	296	1.806648082	4.094344562	9.126958763	8.431635303	6.802394763	7.615791072	5.236441963
7	107	579850	6695899	sprbir	nice	camsed	220	1.903598951	4.248495242	9.200290036	8.338066526	6.738152495	7.999678579	5.295814236
8	108	581636	6694137	sprbir	nice	camsed	325	1.719188776	4.248495242	9.210340372	8.542860938	6.674561392	8.019612794	5.158480421
9	109	585167	6692022	mixdec	nice	camsed	390	1.756132292	4.382026635	9.409191231	8.435983136	6.597145702	7.630461262	5.020585625
10	111	582095	6688877	mixdec	nice	camsed	480	1.556037136	4.787491743	9.332558005	8.171882006	6.398594935	7.455876687	4.54648119
11	112	580702	6686100	sprbir	cloud	camsed	296	1.829376333	4.382026635	9.059517482	8.29654652	7.445417557	7.512071246	5.416988896
12	113	583087	6684132	mixdec	cloud	camsed	360	1.968509981	4.248495242	9.190137665	8.48467	6.265301213	7.820037989	5.178970609
13	114	589716	6679640	spruce	cloud	magm	500	1.93874166	3.912023005	9.047821442	8.218787156	7.580699752	8.080237416	5.26631057
14	115	586913	6675154	birspr	nice	magm	500	1.726331664	4.094344562	8.948975608	8.01301211	8.494538501	8.14322675	5.718671087
15	116	584623	6677522	birspr	nice	magm	556	1.530394705	4.094344562	8.794824928	7.705262475	7.803026644	7.444248649	5.325446034
16	117	589418	6680992	sprprn	nice	magm	500	1.827769907	3.912023005	9.082507	8.154787573	8.050703381	7.605890001	5.799698531
17	118	588861	6672243	spruce	rain	magm	522	1.756132292	4.49980967	9.011889433	8.125630988	8.355379895	7.420578905	5.603593747
18	119	591885	6668839	sprbir	rain	magm	504	1.909542505	4.094344562	8.987196821	8.180320875	8.114623886	7.696212639	6.317706056
19	121	592250	6664455	sprbir	moist	magm	400	1.581038438	4.248495242	8.839276691	8.258422462	7.606884531	7.878534196	5.667809814
20	122	593299	6660953	sprbir	nice	magm	340	1.576914721	4.094344562	9.200290036	7.889833751	8.088254727	8.107720062	5.762994334

Data berisikan 14 variabel dimana 10 diantaranya merupakan variabel numerik yang ingin digunakan. Berikut adalah penjelasan dari masing-masing variabel

No	Nama Variabel	Keterangan
1.	site	Merupakan bentuk unik dari <i>site</i> id yang digunakan
2.	XC	Titik koordinat X ( <i>numeric vector</i> )
3.	YC	Titik koordinat Y ( <i>numeric vector</i> )
4.	forest	<p>Tipe hutan yang terdiri atas</p> <ul style="list-style-type: none"> <li>• “birspr” alias <i>Birch-Spruce</i>,</li> <li>• “mixdec” alias <i>Mixed Deciduous</i> alias hutan campuran,</li> <li>• “pine” alias hutan pinus,</li> <li>• “sprpin” alias <i>Spurce-Pine</i>, dan</li> <li>• “spruce” alias hutan konifer.</li> </ul>
5.	weather	<p>Tipe cuaca yang terdiri atas</p> <ul style="list-style-type: none"> <li>• “cloud” alias berawan,</li> <li>• “moist” alias lembab</li> <li>• “nice” alias bagus atau dalam keadaan normal</li> <li>• “rain” alias hujan</li> </ul>
6.	litho	<p><i>Lithological type</i> alias karakteristik tanah pada lokasi tertentu yang dibagi menjadi beberapa lokasi di daerah sekitar Oslo, Norwegia seperti</p> <ul style="list-style-type: none"> <li>• “camsed” yaitu sedimen Cambro-Silurian</li> <li>• “gneis_o” yaitu <i>gneiss</i> Precambrian-Oslo</li> <li>• “gneis_r” yaitu <i>gneiss</i> Precambrian-Randsfjord</li> <li>• “magm” yaitu batuan magma</li> </ul>
7.	altitude	Ketinggian dari spesies tanaman yang diteliti ( <i>numeric vector</i> )
8.	Cu	<i>Copper</i> alias tembaga ( <i>numeric vector</i> )
9.	Fe	<i>Iron</i> alias besi ( <i>numeric vector</i> )
10.	K	<i>Potassium</i> alias kalium ( <i>numeric vector</i> )
11.	Mg	Magnesium ( <i>numeric vector</i> )
12.	Mn	<i>Manganese</i> alias mangan ( <i>numeric vector</i> )
13.	P	<i>Phosphorus</i> alias fosfor ( <i>numeric vector</i> )
14.	Zn	<i>Zinc</i> alias seng ( <i>numeric vector</i> )

## BAB II PRE-PROCESSING DAN ANALISIS DESKRIPTIF

Pertama-tama akan di-input beberapa *library* yang membantu proses analisis dan pembuatan plot serta input data yang ada. Berikut adalah *codes* dari *input library* dan data serta cuplikan 6 baris pertama dari data

```
> ## Menginput data dan lib
> library(dplyr)
> library(ggplot2)
> library(ggpubr)
> data<-read.csv("C:/Users/Vanny/Documents/UI/KULIAH/S4/ANMUL/FINAL/Oslo.csv")
> ##Melihat data awal
> head(data)
  rownames site      XC      YC forest weather litho altitude      Cu
1         2  102 582865 6706649 sprbir    nice gneis_r    431 1.893112
2         3  103 581155 6720326 sprbir    rain gneis_r    430 1.637053
3         4  104 580482 6711512 sprbir    moist gneis_r    475 1.591274
4         5  105 584618 6701621 spruce    moist gneis_r    410 1.819699
5         6  106 581453 6698513 birspr    nice  camed    296 1.806648
6         7  107 579850 6695899 sprbir    nice  camed    220 1.903599
  Fe      K      Mg      Mn      P      Zn
1 3.912023 9.314700 8.202482 7.419980 7.820038 5.693732
2 3.912023 9.179881 8.045588 7.857094 7.820038 5.257495
3 3.688879 8.895630 8.328451 7.801391 7.889834 5.614222
4 4.248495 9.200290 8.556414 6.645091 7.620705 4.599152
5 4.094345 9.126959 8.431635 6.802395 7.615791 5.236442
6 4.248495 9.200290 8.338067 6.738152 7.999679 5.295814
```

Pada penelitian ini akan dibuang kolom “rownames”, “site”, “XC”, dan “YC” karena tidak dibutuhkan atau tidak berguna pada penelitian. Data baru tanpa variabel yang tidak diinginkan kemudian disimpan pada R dengan nama “data1”. Berikut adalah tipe dari masing-masing kolom data setelah beberapa kolom dibuang

```
> ##Menghapus data yang tidak dibutuhkan
> data1<-data[, -c(1, 2, 3, 4)]
> ##Melihat informasi data
> str(data1)
'data.frame':  332 obs. of  11 variables:
 $ forest : chr  "sprbir" "sprbir" "sprbir" "spruce" ...
 $ weather : chr  "nice" "rain" "moist" "moist" ...
 $ litho   : chr  "gneis_r" "gneis_r" "gneis_r" "gneis_r" ...
 $ altitude: int   431 430 475 410 296 220 325 390 480 296 ...
 $ Cu      : num  1.89 1.64 1.59 1.82 1.81 ...
 $ Fe      : num  3.91 3.91 3.69 4.25 4.09 ...
 $ K       : num  9.31 9.18 8.9 9.2 9.13 ...
 $ Mg      : num  8.2 8.05 8.33 8.56 8.43 ...
 $ Mn      : num  7.42 7.86 7.8 6.65 6.8 ...
 $ P       : num  7.82 7.82 7.89 7.62 7.62 ...
 $ Zn      : num  5.69 5.26 5.61 4.6 5.24 ...
```

Terlihat bahwa kolom “altitude” masih bertipe *integer* sehingga perlu diganti menjadi nilai numerik seperti kolom variabel lainnya dengan *codes* berikut.

```
##Mengubah data integer menjadi numerik
data1$altitude <- as.numeric(data1$altitude)
```

Selanjutnya dengan fungsi *summary* akan dilihat statistik deskriptif dari masing-masing kolom yang berisikan *length*, *class*, dan *mode* untuk variabel kualitatif. Sedangkan untuk

variabel kuantitatif dapat terlihat nilai minimum, nilai quartil, nilai rata-rata, dan nilai maksimum yang terdapat pada gambar berikut

```
> summary(data1)
```

forest	weather	litho	altitude	Cu	Fe
Length:332	Length:332	Length:332	Min. : 30.0	Min. :0.1133	Min. :1.609
Class :character	Class :character	Class :character	1st Qu.:180.0	1st Qu.:0.9478	1st Qu.:2.996
Mode :character	Mode :character	Mode :character	Median :296.0	Median :1.5831	Median :4.094
			Mean :300.4	Mean :1.4114	Mean :3.782
			3rd Qu.:422.5	3rd Qu.:1.7804	3rd Qu.:4.500
			Max. :556.0	Max. :3.2324	Max. :7.576

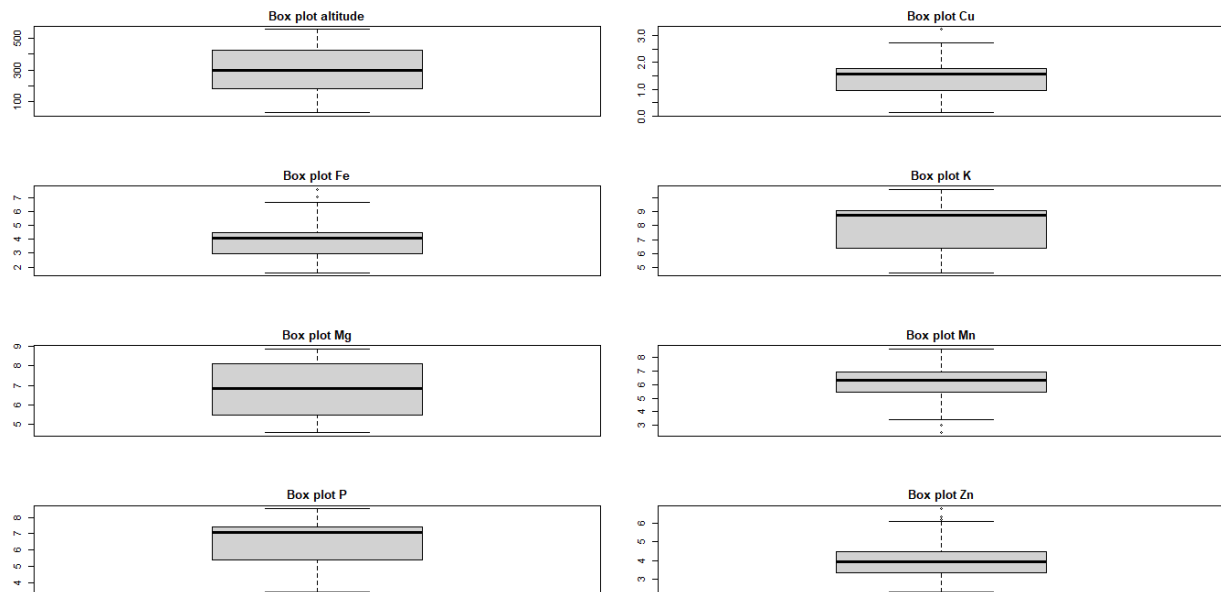
  

K	Mg	Mn	P	Zn
Min. : 4.605	Min. :4.605	Min. :2.485	Min. :3.401	Min. :2.303
1st Qu.: 6.397	1st Qu.:5.481	1st Qu.:5.469	1st Qu.:5.394	1st Qu.:3.353
Median : 8.732	Median :6.862	Median :6.356	Median :7.102	Median :3.922
Mean : 8.009	Mean :6.789	Mean :6.187	Mean :6.493	Mean :3.988
3rd Qu.: 9.060	3rd Qu.:8.082	3rd Qu.:6.924	3rd Qu.:7.409	3rd Qu.:4.447
Max. :10.629	Max. :8.848	Max. :8.629	Max. :8.531	Max. :6.761

Terlihat bahwa di setiap variabel kualitatif memiliki *length*, *class*, dan *mode* yang sama. Pada variabel “altitude” nilai minimum, quartil, rata-rata, dan maksimum yang ada lebih besar sekitar 10 kali lipat dari masing-masing variabel kuantitatif lainnya.

Masuk pada bagian *pre-processing*, akan dilihat apakah data memiliki *missing value* dan *outlier*. Berikut adalah hasil dari jumlah *missing value* dan *boxplot* untuk melihat *outlier*

```
> ##Melihat apakah ada missing value
> sum(is.na(data1))
[1] 0
```



Terlihat bahwa tidak ada *missing value* sehingga data tidak perlu dimodifikasi. Untuk *boxplot* pada masing-masing variabel kuantitatif terlihat beberapa variabel seperti “Cu”, “Fe”, dan “Zn” memiliki *outlier* yang tidak terlalu banyak sehingga *outlier* ini di diamankan saja. Tidak dilakukan

penghapusan pada *outlier* karena ditakutkan nilai *distance* dapat berubah signifikan seiring penghapusan nilai tertentu sehingga kevalidan dari nilai tersebut diragukan.

### BAB III METODE

Masalah yang akan dibahas pada penelitian adalah penggunaan *Multidimensional Scaling* (MDS) untuk mengolah data agar didapatkan peta persepsi dari informasi komposisi konsentrasi bahan kimia pada spesies tanaman di Oslo, Norwegia dengan berbagai faktor seperti kondisi cuaca, tipe hutan, dan jenis tanah tanaman. Terdapat tiga objek yang akan diteliti yaitu kondisi cuaca, tipe hutan, dan jenis tanah dengan variabel ukur yaitu titik koordinat X, titik koordinat Y, *altitude*, kadar Cu, Fe, K, Mg, Mn, P, dan Zn.

Analisis *Multidimensional Scaling* (MDS) adalah salah satu teknik multivariat yang dapat digunakan untuk menentukan posisi suatu objek relatif terhadap objek lainnya berdasarkan kemiripannya. Kegunaan dari analisis ini antara lain untuk mendapatkan posisi relatif suatu objek dibandingkan objek yang lain, mengelompokkan objek sebagai salah satu alternatif dari *cluster analysis*. Hal ini selaras dengan tujuan awal penelitian untuk mengelompokkan variabel ukur berdasarkan objek yang ada dengan *output* grafis pada suatu bidang multidimensi yang dapat membantu mengidentifikasi informasi yang ada.

Dalam pelaksanaannya, *Multidimensional Scaling* (MDS) akan menganalisis data berdasarkan tingkat kemiripan dari variabel ukur antar objek. Berdasarkan tipe atau skala yang digunakan, MDS dibagi menjadi dua tipe yaitu

a) *Multidimensional Scaling* Metrik

*Multidimensional Scaling Metric* dapat digunakan untuk data jarak yang ditaksir dalam skala interval atau rasio. Pada *Multidimensional Scaling Metric* asumsi data *input* merupakan jarak yang sebenarnya atau tidak diabaikan karena pada tipe ini hanya akan menyusun bentuk geometri dari titik objek sedekat mungkin dari *input* jarak yang diberikan

b) *Multidimensional Scaling* non-Metrik

*Multidimensional Scaling non-Metric* dapat digunakan untuk data jarak yang ditaksir dalam skala nominal atau ordinal. Pada *Multidimensional Scaling non-Metric*, data jarak yang digunakan dalam transformasi bernilai sama dengan data yang sebenarnya sehingga dapat dilakukan operasi aritmatika terhadap nilai ketidaksamaannya.



Berikut adalah tahapan dalam melakukan Analisis *Multidimensional Scaling* (MDS)

1. Menghitung matriks jarak

Perhatikan bahwa *original distance*  $\delta_{ij}$  yaitu hasil asli dari jarak antar observasi  $y_i$  dan  $y_j$  pada dimensi  $p$ . Jarak  $\delta_{ij}$  didasari oleh aproksimasi kedekatan berdasarkan penilaian manusia. Sedangkan, *final distance*  $d_{ij}$  yang mewakili  $n$  item pada koordinat dimensi yang lebih rendah sehingga  $d_{ij} \cong \delta_{ij}$  untuk semua  $i, j$ .  $d_{ij}$  biasanya merupakan jarak Euclidean. Berikut adalah rumus dari *original distance*

$$\delta_{ij} = [(y_i - y_j)'(y_i - y_j)]^{1/2}$$

Untuk melakukan *metric multidimensional scaling* atau *classical solution*. Akan dibuat matriks  $\mathbf{D} = (\delta_{ij})$  berukuran  $n \times n$  dengan mencari sebanyak  $n$  point pada  $k$  dimensi sehingga jarak antar titik  $d_{ij} \cong \delta_{ij}$ . Biasanya dalam membuat sebuah *plot* digunakan nilai  $k = 2$ .

2. Membuat matriks  $\mathbf{A}$

$$\mathbf{A} = (a_{ij}) = (-\frac{1}{2}\delta_{ij}^2) \text{ berukuran } n \times n$$

3. Membuat matriks  $\mathbf{B}$

$$\mathbf{B} = (b_{ij}) \text{ dimana}$$

- $b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}$
- $\bar{a}_{i.} = \sum_{j=1}^n a_{ij}/n$
- $\bar{a}_{.j} = \sum_{i=1}^n a_{ij}/n$
- $\bar{a}_{..} = \sum_{ij} a_{ij}/n^2$

Sehingga matriks  $\mathbf{B}$  dapat ditulis dengan persamaan

$$\mathbf{B} = (\mathbf{I} - \frac{1}{n}\mathbf{J}) \mathbf{A} (\mathbf{I} - \frac{1}{n}\mathbf{J})$$

4. Mencari nilai *eigen value* dan *eigen vector*

Dengan **B** yang merupakan matriks yang simetris maka dapat dicari nilai *eigen values* dan *eigen vector* dengan persamaan

$$\det(B - \lambda I) \text{ dan } \det(B - \lambda I)X$$

Dimana matriks **X** adalah nilai *eigen vector*

5. Menghitung jarak koordinat yang terbentuk

Berdasarkan *eigen vector* matriks **X** akan dicari jarak Euclidean dari koordinat yang terbentuk yang dilambangkan dengan  $\hat{D}$ . Jarak Euclidean tersebut terbentuk dari persamaan  $d_{ij} = (x_i - x_j)'(x_i - x_j)$

6. Menghitung nilai STRESS

Nilai STRESS yang kecil menerangkan hubungan monoton yang terbentuk antara ketidaksamaan dengan *disparities* dan kriteria peta persepsi yang terbentuk semakin sempurna. Berikut adalah rumus dari STRESS

$$S = \frac{\sum_{i=j}^n (d_{ij} - \hat{d}_{ij})^2}{\sum_{i=j}^n d_{ij}^2}$$

Pada penelitian ini akan digunakan *Metric Multidimensional Scaling* dikarenakan beberapa pertimbangan berikut

- Berdasarkan contoh-contoh dari buku dan jurnal, penggunaan *Metric Multidimensional Scaling* lebih umum digunakan.
- *Metric* MDS dapat menempatkan objek dalam ruang berdimensi rendah sehingga jarak sesungguhnya sesuai dengan jarak pada matriks Euclidean
- *Metric* MDS mempertahankan akurasi metrik dalam perhitungan jarak antar objek
- *Metric* MDS memberikan kualitas representasi jarak metrik yang jelas

## BAB IV PENGOLAHAN DATA ANALISIS DAN HASIL

Pertama-tama akan dicari matriks *distance* dengan metode Euclidean, namun kita perlu membuat modifikasi dari “data1” bernama “data1t” yang merupakan hasil *transpose* dari “data1”. Euclidean *distance* dapat diterapkan pada variabel kuantitatif saja sehingga kita membuat subset dari “data1t” yang hanya berisikan variabel kuantitatif. Variabel kuantitatif yang digunakan antara lain adalah “altitude”, “Cu”, “Fe”, “K”, “Mg”, “Mn”, “P”, dan “Zn”. Berikut adalah *codes transpose* dan pencarian *Euclidean distance* menggunakan modul R.

```
##Membuat data transpose untuk distance dari data
data1t<-t(data1)

##MDS jenis Hutan dan Cuaca
###Membuat subset data ukur
data_ukur <- data1t[c("altitude", "Cu", "Fe", "K", "Mg", "Mn", "P", "Zn"), ]

> distance_matrix <- as.matrix(distance_matrix);distance_matrix
```

	altitude	Cu	Fe	K	Mg	Mn	P	Zn
altitude	0.000	6028.38710	5990.47712	5919.89340	5940.28879	5948.95960	5944.61258	5986.01266
Cu	6028.387	0.00000	45.73638	122.63294	99.93270	88.80387	94.57514	49.21574
Fe	5990.477	45.73638	0.00000	79.61290	57.68713	49.36396	51.85849	23.71607
K	5919.893	122.63294	79.61290	0.00000	24.78654	42.42799	30.06943	79.04730
Mg	5940.289	99.93270	57.68713	24.78654	0.00000	24.96674	13.51839	57.01767
Mn	5948.960	88.80387	49.36396	42.42799	24.96674	0.00000	20.96953	44.28749
P	5944.613	94.57514	51.85849	30.06943	13.51839	20.96953	0.00000	51.22459
Zn	5986.013	49.21574	23.71607	79.04730	57.01767	44.28749	51.22459	0.00000

Selanjutnya kita akan melihat apakah *original distance* memiliki nilai yang mirip dengan *final distance* menggunakan Euclidean. Hal tersebut dilakukan dengan mencari matriks **A**, **B**, dan **Z**. Berikut adalah *codes* dan hasil dari matrik-matriks tersebut

```
##Mencari matriks
A<- -(1/2)*distance_matrix^2;A
n<-nrow(distance_matrix);n
I<-diag(n)
J<-matrix(rep(1, times=n^2), ncol=n)
B<-(I-(1/n)*J)%*%A%*%(I-(1/n)*J);B

#Mencari eig val dan eig vec
v<-eigen(B)$vec;v
lambda<-diag(eigen(B)$values);lambda
v1<-v[,1:2]
lambda1<-diag(eigen(B)$values[1:2])
Z<-v1%*%sqrt(lambda1)
as.matrix(dist(Z, diag=TRUE))
distance_matrix
```

```
> A<- -(1/2)*distance_matrix^2;A
```

	altitude	Cu	Fe	K	Mg	Mn	P	Zn
altitude	0	-18170725.534	-1.794291e+07	-1.752257e+07	-1.764352e+07	-1.769506e+07	-1.766921e+07	-1.791617e+07
Cu	-18170726	0.000	-1.045908e+03	-7.519419e+03	-4.993272e+03	-3.943064e+03	-4.472229e+03	-1.211095e+03
Fe	-17942908	-1045.908	0.000000e+00	-3.169107e+03	-1.663902e+03	-1.218400e+03	-1.344652e+03	-2.812259e+02
K	-17522569	-7519.419	-3.169107e+03	0.000000e+00	-3.071862e+02	-9.000672e+02	-4.520855e+02	-3.124238e+03
Mg	-17643515	-4993.272	-1.663902e+03	-3.071862e+02	0.000000e+00	-3.116690e+02	-9.137337e+01	-1.625507e+03
Mn	-17695060	-3943.064	-1.218400e+03	-9.000672e+02	-3.116690e+02	0.000000e+00	-2.198605e+02	-9.806909e+02
P	-17669209	-4472.229	-1.344652e+03	-4.520855e+02	-9.137337e+01	-2.198605e+02	0.000000e+00	-1.311979e+03
Zn	-17916174	-1211.095	-2.812259e+02	-3.124238e+03	-1.625507e+03	-9.806909e+02	-1.311979e+03	0.000000e+00

```

> B<-(I-(1/n)*J)%*%A%*%(I-(1/n)*J);B
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
[1,] 27246279 -4220227.4 -4022694.9 -3654054.5 -3760692.6 -3805971.6 -3783312.4 -3999325.9
[2,] -4220227  654716.7  623385.9  565213.6  582048.2  589364.1  585643.4  619855.4
[3,] -4022695  623385.9  594146.9  539279.0  555092.7  561803.9  558486.1  590500.4
[4,] -3654055  565213.6  539279.0  490749.4  504750.6  510423.4  507679.9  535958.6
[5,] -3760693  582048.2  555092.7  504750.6  519366.2  525320.2  522349.0  551765.7
[6,] -3805972  589364.1  561803.9  510423.4  525320.2  531897.6  528486.2  558676.2
[7,] -3783312  585643.4  558486.1  507679.9  522349.0  528486.2  525514.5  555153.4
[8,] -3999326  619855.4  590500.4  535958.6  551765.7  558676.2  555153.4  587416.2

> v<-eigen(B)$vec;v
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
[1,] 0.9352879 -0.01511658 -0.002816711 7.937394e-05 -6.662154e-05 -0.0002589343 -0.0002072084 -0.3535534
[2,] -0.1448693 -0.64809905 -0.275625573 7.108037e-03 -4.678359e-01 0.3646346471 0.0783050000 -0.3535534
[3,] -0.1380880 -0.18710818 -0.456520433 -3.988904e-01 4.898826e-01 -0.4555220572 0.0768925054 -0.3535534
[4,] -0.1254327 0.53511878 -0.203393016 1.912626e-01 -5.857564e-02 0.1803542664 0.6774995461 -0.3535534
[5,] -0.1290936 0.32177267 -0.119214756 1.672976e-01 -5.176472e-01 -0.4804026699 -0.4624405067 -0.3535534
[6,] -0.1306482 0.05656364 0.670196981 -6.106192e-01 -1.473797e-01 0.0128870385 0.1040430174 -0.3535534
[7,] -0.1298702 0.24197380 -0.066925714 8.675412e-03 3.816097e-01 0.5919579873 -0.5467978839 -0.3535534
[8,] -0.1372859 -0.30510508 0.454299221 6.350866e-01 3.200128e-01 -0.2136502779 0.0727055302 -0.3535534

> Z<-as.matrix(dist(Z, diag=TRUE));Z
      1      2      3      4      5      6      7      8
1  0.000 6028.38214 5990.46508 5919.89005 5940.284014 5948.93902 5944.608760 5985.99921
2 6028.382 0.00000 43.85394 122.47586 99.620787 86.29065 94.005931 45.41931
3 5990.465 43.85394 0.00000 78.69660 55.838251 43.14123 50.286234 7.22476
4 5919.890 122.47586 78.69660 0.00000 22.859476 37.09695 28.491874 77.50294
5 5940.284 99.62079 55.83825 22.85948 0.000000 15.41854 5.787173 54.75476
6 5948.939 86.29065 43.14123 37.09695 15.418541 0.00000 9.912390 40.92008
7 5944.609 94.00593 50.28623 28.49187 5.787173 9.91239 0.000000 49.03246
8 5985.999 45.41931 7.22476 77.50294 54.754755 40.92008 49.032461 0.00000

```

Terlihat bahwa matrik  $Z$  memiliki nilai yang hampir serupa dengan matriks *distance*. Sehingga kita dapat menggunakan *Metric Multidimensional Scaling* (MDS).

Untuk mendapatkan visualisasi MDS pertama-tama perlu membuat fungsi dari MDS terlebih dahulu yang berasal dari matriks *distance* serta memilih dimensi  $k$  yaitu 2. Setelah struktur MDS awal sudah ada maka perlu kita tambahkan nilai dari variabel kualitatif yang ingin dibentuk *plot* MDS-nya. Berikut adalah *codes* untuk fungsi MDS dengan bantuan *modul* R.

```

###MDS Codes
mds_result <- cmdscale(distance_matrix, k = 2)

###Mengconvert hasil MDS menjadi data frame
mds_df <- as.data.frame(mds_result)
colnames(mds_df) <- c("Dim1", "Dim2")

###Menambahkan variabel kategorik ke data frame hasil MDS
mds_df$jenis_hutan <- data1$forest
mds_df$cuaca <- data1$weather
mds_df$tanah <- data1$litho

```

```

###visualisasikan hasil Jenis Hutan dan Cuaca
ggplot(mds_df, aes(x = Dim1, y = Dim2)) +
  geom_point(aes(color = jenis_hutan, shape = cuaca), size = 3) +
  labs(title = "Visualisasi MDS Berdasarkan Jenis Hutan dan Cuaca",
        x = "Dimensi 1", y = "Dimensi 2") +
  theme_minimal()

#MDS Jenis Hutan
ggplot(mds_df, aes(x = Dim1, y = Dim2)) +
  geom_point(aes(color = jenis_hutan), size = 3) +
  labs(title = "Visualisasi MDS Berdasarkan Jenis Hutan",
        x = "Dimensi 1", y = "Dimensi 2") +
  theme_minimal()

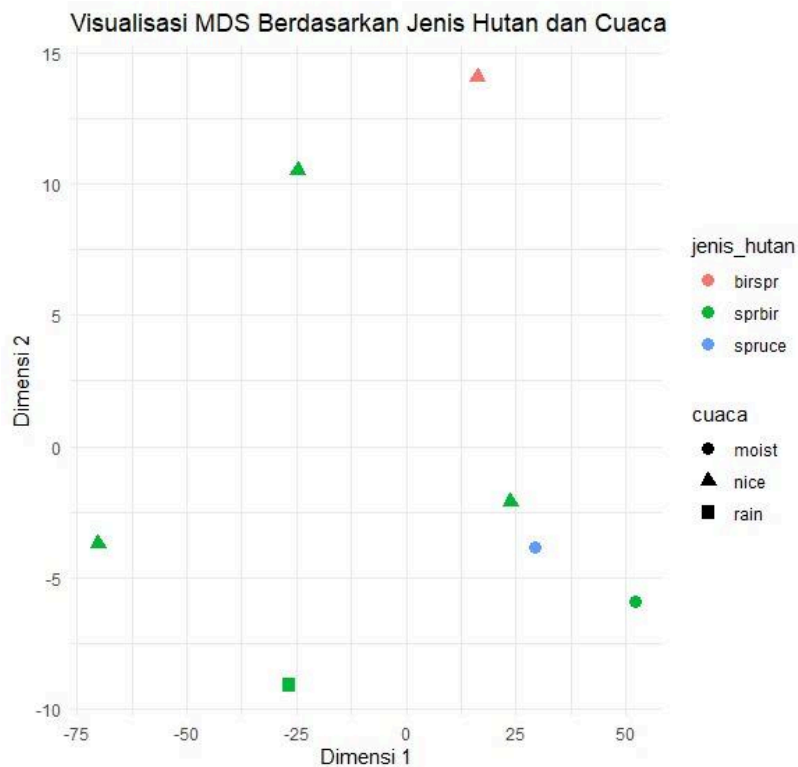
#MDS Jenis Cuaca
ggplot(mds_df, aes(x = Dim1, y = Dim2)) +
  geom_point(aes(colour = cuaca), size = 3) +
  labs(title = "Visualisasi MDS Berdasarkan Cuaca",
        x = "Dimensi 1", y = "Dimensi 2") +
  theme_minimal()

#MDS Jenis Tanah
ggplot(mds_df, aes(x = Dim1, y = Dim2)) +
  geom_point(aes(colour = tanah), size = 3) +
  labs(title = "Visualisasi MDS Berdasarkan Tanah",
        x = "Dimensi 1", y = "Dimensi 2") +
  theme_minimal()

```

Pada penelitian ini, akan dihasilkan beberapa *plot* hasil MDS yaitu

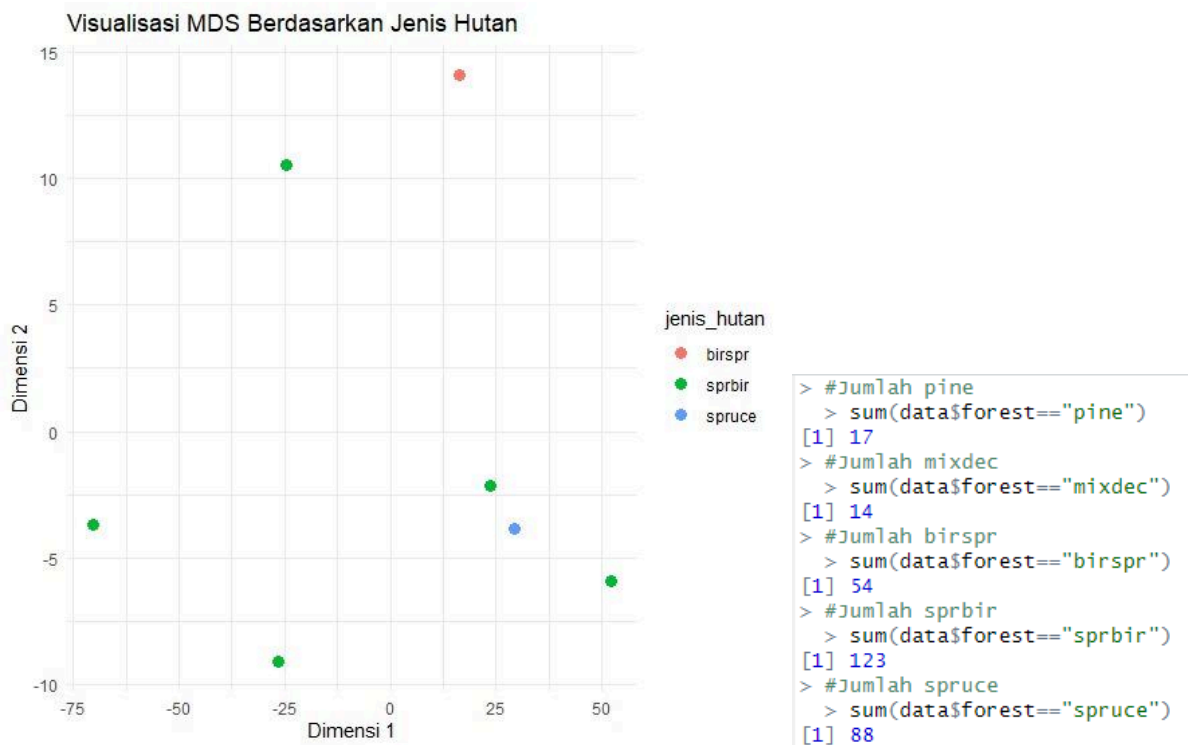
1. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Jenis Hutan dan Cuaca



Berikut adalah interpretasi dari visualisasi MDS di atas

- Setiap titik mempresentasikan sebuah observasi dengan jenis hutan dan cuaca tertentu.
- Spesies tanaman dibagi berdasarkan jenis hutan dengan ditandai oleh warna merah untuk “birspr”, warna hijau untuk “sprbir”, dan warna biru untuk “spruce”. Tipe hutan “mixdec” dan “pine” tidak muncul pada plot ini dikarenakan jumlah kombinasi jenis hutan dengan cuaca yang tergolong sedikit.
- Spesies tanaman dibagi berdasarkan cuaca dengan ditandai oleh lingkaran untuk “moist”, segitiga untuk “nice”, dan kotak untuk “rain”. Tipe hutan “cloud” tidak muncul pada plot ini dikarenakan jumlah kombinasi jenis hutan dengan cuaca yang tergolong sedikit.
- Spesies tanaman pada jenis hutan “sprbir” dengan cuaca “nice” menunjukkan karakteristik serupa yang kuat dengan spesies tanaman pada jenis hutan “spruce” dengan cuaca “moist” berdasarkan jarak pada dimensi 1.
- Spesies tanaman pada jenis hutan “birspr” dengan cuaca “nice” berada pada atas plot dan letaknya cukup jauh dari titik lainnya, menunjukkan karakteristik yang berbeda dibandingkan jenis hutan dan cuaca species tanaman lainnya.
- Spesies tanaman pada jenis hutan “sprbir” tersebar di berbagai wilayah yang menunjukkan variabilitas tinggi.

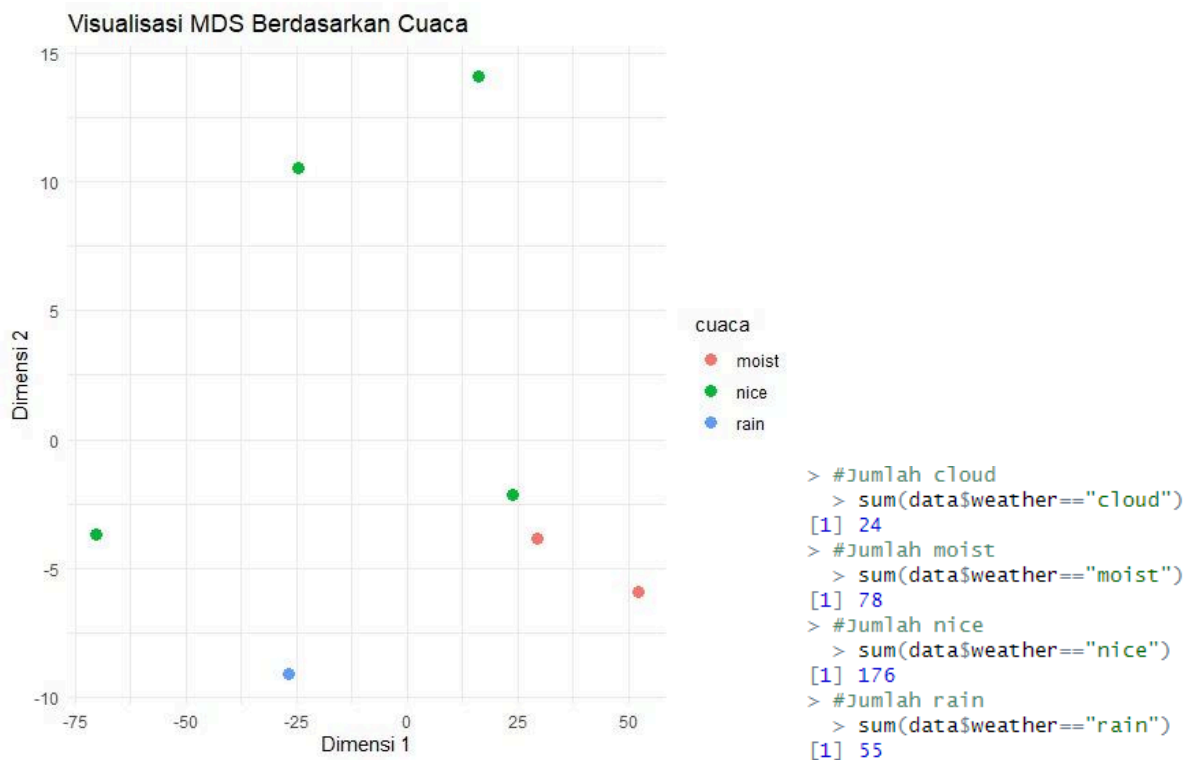
## 2. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Jenis Hutan



Berikut adalah interpretasi dari visualisasi MDS di atas

- Setiap titik mempresentasikan sebuah observasi dengan jenis hutan yang berbeda-beda
- Spesies tanaman dibagi berdasarkan jenis hutan dengan ditandai oleh warna merah untuk “birspr”, warna hijau untuk “sprbir”, dan warna biru untuk “spruce”. Tipe hutan “mixdec” dan “pine” tidak muncul pada plot ini dikarenakan jumlah jenis hutan yang tergolong sedikit dibandingkan jenis hutan yang lain.
- Spesies tanaman pada jenis hutan “sprbir” menunjukkan karakteristik serupa yang kuat dengan spesies tanaman pada jenis hutan “spruce” berdasarkan jarak pada dimensi 1.
- Spesies tanaman pada jenis hutan “birspr” berada pada atas plot dan letaknya cukup jauh dari titik lainnya, hal ini terjadi diduga karena banyak nilai dari “birspr” yang juga cukup kecil dibandingkan spesies tanaman pada jenis hutan lainnya.
- Spesies tanaman pada jenis hutan “sprbir” tersebar di berbagai wilayah yang menunjukkan variabilitas paling tinggi.
- Spesies tanaman pada jenis hutan “spruce” berada pada tengah titik koordinat, hal ini terjadi diduga karena banyak nilai dari “spruce” yang berada di tengah-tengah tidak terlalu sedikit dan tidak terlalu banyak.

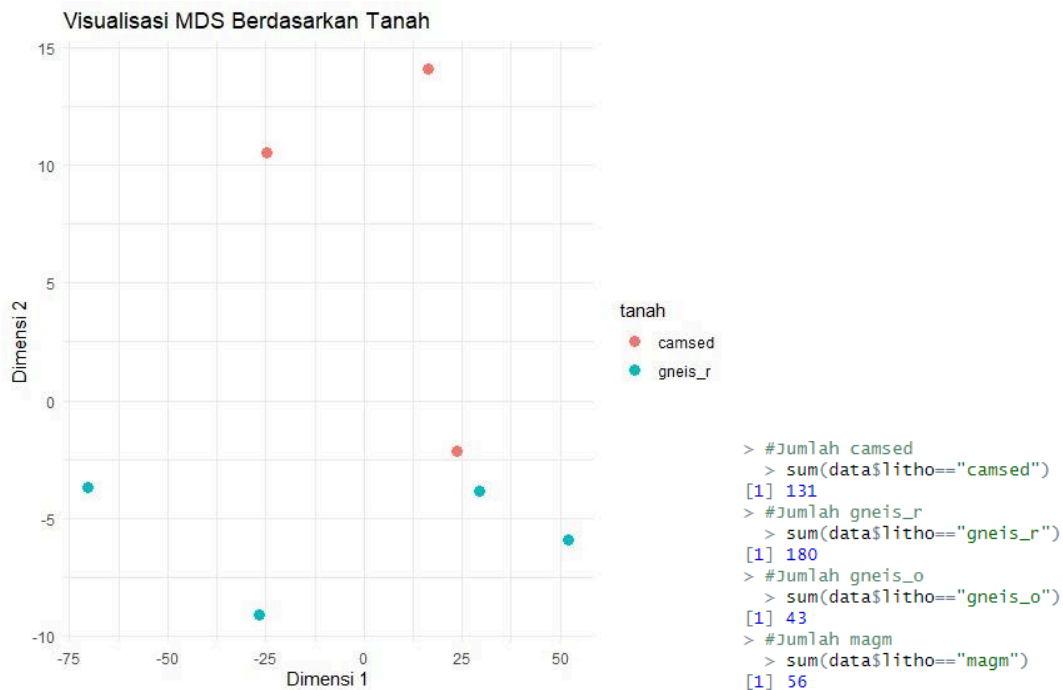
### 3. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Cuaca



Berikut adalah interpretasi dari visualisasi MDS di atas

- Setiap titik mempresentasikan sebuah observasi dengan cuaca yang berbeda-beda
- Spesies tanaman dibagi berdasarkan cuaca dengan ditandai oleh merah untuk “moist”, hijau untuk “nice”, dan biru untuk “rain”. Cuaca “cloud” tidak muncul pada plot ini dikarenakan jumlahnya yang tergolong sedikit.
- Spesies tanaman pada cuaca “rain” berada pada bawah plot dan letaknya tidak mendekati titik lainnya, hal ini diduga terjadi karena nilai “rain” yang juga cukup kecil dibanding banyak nilai cuaca lainnya.
- Spesies tanaman pada cuaca “nice” tersebar di berbagai wilayah yang menunjukkan variabilitas paling tinggi.
- Spesies tanaman pada cuaca “moist” berada pada tengah titik koordinat, hal ini terjadi diduga karena banyak nilai dari “nice” yang berada di tengah-tengah tidak terlalu sedikit dan tidak terlalu banyak.

#### 4. Visualisasi MDS Konsentrasi Bahan Kimia pada Tumbuhan terhadap Tanah Hidup



Berikut adalah interpretasi dari visualisasi MDS di atas

- Setiap titik mempresentasikan sebuah observasi dengan karakteristik tanah yang berbeda-beda
- Spesies tanaman dibagi berdasarkan cuaca dengan ditandai oleh merah untuk “camsed”, biru untuk “gneis\_r”. Karakteristik tanah “gneis\_o” dan “magm” tidak muncul pada plot ini diduga dikarenakan jumlahnya yang tergolong sedikit.
- Spesies tanaman pada tanah “camsed” menyebar pada bagian atas plot sedangkan “gneis\_r” tersebar pada bagian bawah plot yang mengindikasikan perbedaan karakteristik dari keduanya.



## 5. Nilai STRESS

Setelah memperoleh plot dari MDS, untuk mengetahui apakah plot tersebut sudah baik dalam menerangkannya maka dicarilah nilai STRESS dari fungsi MDS. Didapatkan nilai 0,089 dimana menurut sumber bacaan apabila mengambil nilai  $k = 2$  plot tersebut baik dalam mempresentasikan hasil jika nilai STRESS di bawah 0,2.

```
> stress <- sum((dist(mds_df) - distance_matrix)^2) / sum(distance_matrix^2)
warning message:
In dist(mds_df) : NAs introduced by coercion
> stress
[1] 0.08907517
```

## BAB V PENUTUP

Dalam penelitian yang mengeksplorasi kemiripan dan perbedaan konsentrasi bahan kimia pada spesies tanaman di sekitar kota Oslo, Norwegia, dengan mempertimbangkan faktor lingkungan seperti jenis hutan, cuaca, dan jenis tanah. Variabel ukur yang mengamati faktor lingkungan pada penelitian adalah “altitude”, “Cu”, “Fe”, “K”, “Mg”, “Mn”, “P”, dan “Zn”. Metode Multidimensional Scaling (MDS) dipilih untuk menganalisis data dari 332 observasi tanaman yang diambil dalam radius 120 km mengelilingi Oslo.

Beberapa temuan utama dari penelitian menggunakan MDS adalah sebagai berikut:

1. Pengaruh Jenis Hutan dan Cuaca terhadap Konsentrasi Bahan Kimia
  - Terdapat kemiripan yang cukup jelas berdasarkan jenis hutan dan cuaca tertentu. Tanaman dari hutan "sprbir" dengan cuaca "nice" dan tanaman hutan "spruce" dengan cuaca "moist" menunjukkan karakteristik serupa berdasarkan jarak pada dimensi 1.
  - Tanaman dari hutan "birspr" dengan cuaca "nice" berada di posisi yang berbeda signifikan dibandingkan dengan kombinasi jenis hutan dan cuaca lainnya, hal ini menunjukkan karakteristik yang unik pada tanaman hutan “birspr”.
2. Pengaruh Jenis Hutan
  - Tanaman dari hutan "sprbir" memiliki variabilitas tertinggi dalam karakteristik kimia.
  - Tanaman dari hutan "spruce" berada di tengah-tengah plot, menunjukkan variabilitas dan konsentrasi bahan kimia yang moderat.
3. Pengaruh Cuaca
  - Tanaman yang tumbuh dalam kondisi cuaca "rain" menunjukkan karakteristik yang berbeda dengan nilai konsentrasi bahan kimia yang lebih rendah.
  - Tanaman dalam kondisi cuaca "nice" menunjukkan variabilitas tertinggi dalam konsentrasi bahan kimia.
  - Tanaman dalam kondisi cuaca "moist" menunjukkan nilai konsentrasi yang moderat karena berada di tengah-tengah plot.
4. Pengaruh Jenis Tanah
  - Tanaman yang tumbuh di tanah "camsed" cenderung berada di bagian atas plot, sedangkan tanaman di tanah "gneis\_r" berada di bagian bawah, menunjukkan perbedaan karakteristik kimia yang jelas antara kedua jenis tanah tersebut.

Analisis MDS pada penelitian ini memberikan visualisasi yang efektif dalam memahami struktur hubungan antar variabel ukur yang ada pada spesies tanaman dengan variabel kategorik berdasarkan kemiripan dengan jarak Euclidean. Hal ini dibuktikan dengan nilai STRESS di

bawah 0,2 yang mengindikasikan bahwa plot dua dimensi pada penelitian sudah cukup baik dalam mempresentasikan hasil analisis data.

## DAFTAR PUSTAKA

Oslo Transect Subset Data. R Documentation.

<https://vincentarelbundock.github.io/Rdatasets/doc/heplots/Oslo.html>

<https://vincentarelbundock.github.io/Rdatasets/csv/heplots/Oslo.html>

Rencher, A. C., & Christensen, W. F. (1995). *Methods of Multivariate Analysis*| Wiley Series in Probability and Statistics.

Kruskal, Joseph. (1977) The Relationship between Multidimensional Scaling and Clustering. In *Classification and clustering* (pp. 1-28). Academic Press.

HMPS Statistika FMIPA UNM. (2023). Multidimensional Scaling (MDS). Universitas Negeri Makassar.

<https://hmpsstatistikafmipaunm.com/2023/12/04/multidimensional-scaling-mds/>

Kassambara. (2017). *Principal Component Methods in R: Practical Guide*| Multidimensional Scaling Essentials: Algorithms and R Code. STHDA.

<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/122-multidimensional-scaling-essentials-algorithms-and-r-code/>

Letten, Andrew. (2016). Multidimensional Scaling| Environmental Computing.

<https://environmentalcomputing.net/graphics/multivariate-vis/mds/>

## LAMPIRAN

Berikut adalah lampiran linimasa yang berisikan data dan codes R

📁 FINAL PROJECT

Data diambil pada 05 Juni 2024