

ProjekAkhir

Kelompok_DS_069_106_115

2024-12-02

Analisis Tingkat Kerusakan Akibat Kebakaran Hutan Menggunakan Algoritma K-Means Clustering

Library

```
library(dslabs)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidymodels)
```

```
## -- Attaching packages ----- tidymodels 1.2.0 --
## v broom       1.0.6      v rsample    1.2.1
## v dials       1.3.0      v tune       1.2.1
## v infer       1.0.7      v workflows  1.1.4
## v modeldata   1.4.0      v workflowsets 1.1.0
## v parsnip     1.2.1      v yardstick  1.3.1
## v recipes     1.1.0
## -- Conflicts ----- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed() masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Use suppressPackageStartupMessages() to eliminate package startup messages
```

```
library(here)
```

```
## here() starts at C:/Users/L E N O V O/OneDrive/Documents/Kuliah/Data Science/Tugas Akhir Praktikum
```

```
library(cluster)
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.4.2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

Data Preparation

Memilah kolom

```
dataset = read.csv(here("AK_CA_Burned_Plot_Data_1983_2016.csv"))
clean_data = dataset %>%
  select(burn_depth, prop_sol_combusted, total_c_combusted, ag_biomass_combusted,
         residual_sol_c, mean_residual_org_layer_depth, initial_spread_index)
colnames(clean_data) = c("kedalaman_pembakaran", "proporsi_lapisan_organik_terbakar",
                        "karbon_terbakar", "biomassa_terbakar", "karbon_residu_tanah",
                        "kedalaman_rata_rata_lapisan_residu", "tingkat_penyebaran_api")
View(dataset)
head(clean_data)
```

```
##   kedalaman_pembakaran proporsi_lapisan_organik_terbakar karbon_terbakar
## 1                    -9999                    -9999                    -9999
## 2                    -9999                    -9999                    -9999
## 3                    -9999                    -9999                    -9999
## 4                    -9999                    -9999                    -9999
## 5                    -9999                    -9999                    -9999
## 6                    -9999                    -9999                    -9999
##   biomassa_terbakar karbon_residu_tanah kedalaman_rata_rata_lapisan_residu
## 1                   0                -9999                        67
## 2                   0                -9999                        75
## 3                   0                -9999                        57
## 4                   0                -9999                        64
## 5                   0                -9999                        71
## 6                   0                -9999                        62
##   tingkat_penyebaran_api
## 1                    -9999
## 2                    -9999
## 3                    -9999
## 4                    -9999
## 5                    -9999
## 6                    -9999
```

Menghapus nilai NA

```
clean_data[clean_data == -9999] = NA

clean_data = na.omit(clean_data)

View(clean_data)
head(clean_data)
```

```
##      kedalaman_pembakaran proporsi_lapisan_organik_terbakar karbon_terbakar
## 200                20.7                0.782                4260.1
## 201                10.2                0.695                3024.3
## 202                20.1                0.652                2423.2
## 203                15.8                0.462                1519.8
## 204                25.4                0.897                2660.1
## 205                18.3                0.680                2815.5
##      biomassa_terbakar karbon_residu_tanah kedalaman_rata_rata_lapisan_residu
## 200            3143.378            1293.2                5.78
## 201            1797.920            1094.0                4.48
## 202            1343.002            3339.4                10.74
## 203             879.180            3455.2                18.42
## 204             989.716            1662.5                 2.91
## 205            1524.107            1565.0                8.62
##      tingkat_penyebaran_api
## 200                3.21
## 201                3.21
## 202                4.07
## 203                4.07
## 204                2.23
## 205                2.31
```

Scaling / Normalisasi data

```
clean_data_scaled = clean_data %>% scale()
head(clean_data_scaled)
```

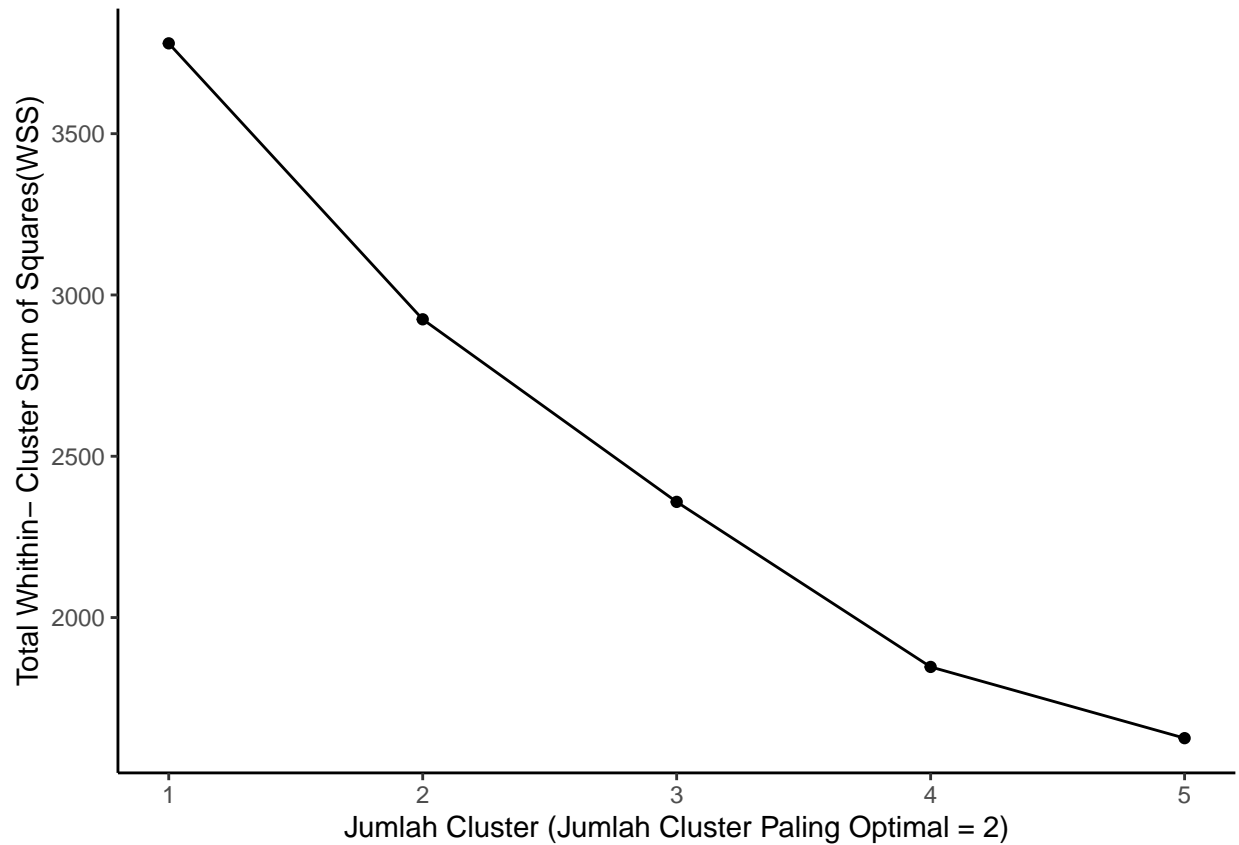
```
##      kedalaman_pembakaran proporsi_lapisan_organik_terbakar karbon_terbakar
## 200            2.04399602            1.0871961            0.5090291
## 201           -0.07551976            0.7576619           -0.1674922
## 202            1.92288083            0.5947887           -0.4965560
## 203            1.05488866           -0.1248837           -0.9911096
## 204            2.99273166            1.5227873           -0.3668684
## 205            1.55953527            0.7008457           -0.2817968
##      biomassa_terbakar karbon_residu_tanah kedalaman_rata_rata_lapisan_residu
## 200            3.2084170           -0.5491574           -0.4886713
## 201            1.4424245           -0.5630046           -0.5282147
## 202            0.8453180           -0.4069186           -0.3377979
## 203            0.2365245           -0.3988689           -0.1041876
## 204            0.3816094           -0.5234861           -0.5759710
## 205            1.0830289           -0.5302636           -0.4022841
##      tingkat_penyebaran_api
## 200           -0.9488640
```

```
## 201          -0.9488640
## 202          -0.7376108
## 203          -0.7376108
## 204          -1.1895944
## 205          -1.1699430
```

Modeling

Menentukan nilai K optimal

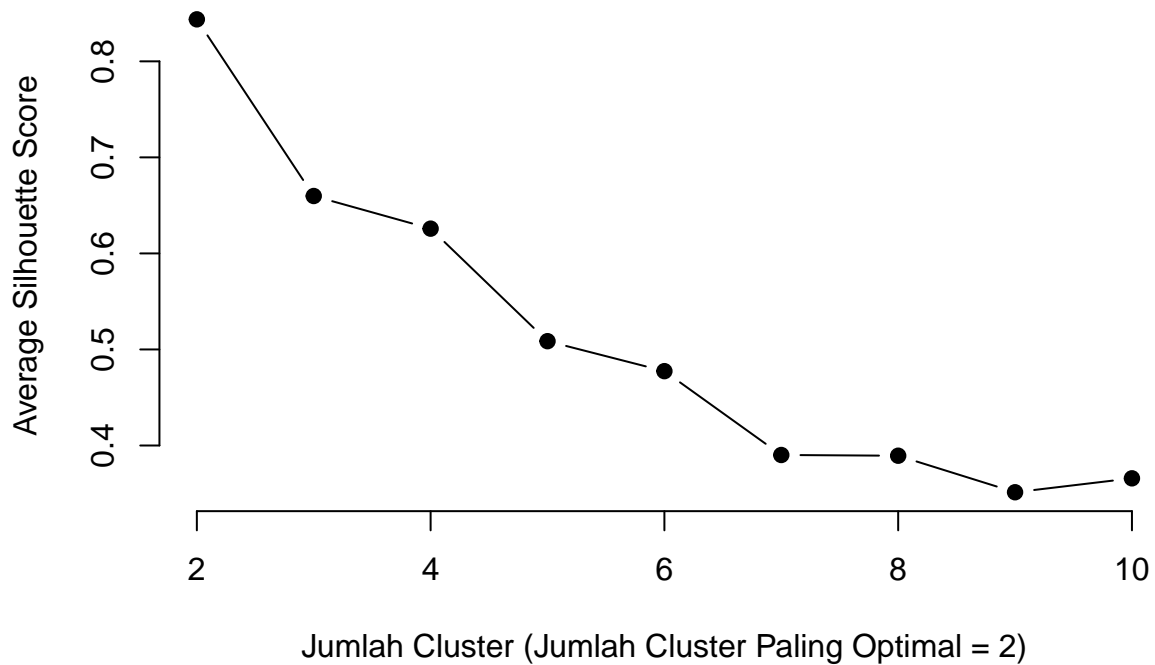
```
wss = sapply(1:5, function(k){
  set.seed(98)
  kmeans(clean_data_scaled,
    centers = k,
    nstart = 5)$tot.withinss
})
elbow_data=data.frame(
  k=1:5,
  wss=wss
)
ggplot(
  elbow_data,
  aes(
    x=k,
    y=wss
  )
) + geom_line() + geom_point() + labs(
  x = "Jumlah Cluster (Jumlah Cluster Paling Optimal = 2)",
  y = "Total Whithin- Cluster Sum of Squares(WSS)"
) + theme_classic()
```



```
silhouette_scores <- numeric()
for (k in 2:10) {
  kmeans_result <- kmeans(clean_data, centers = k, nstart = 25)
  sil <- silhouette(kmeans_result$cluster, dist(clean_data))
  silhouette_scores[k] <- mean(sil[, 3])
}

plot(2:10, silhouette_scores[2:10], type = "b", pch = 19, frame = FALSE,
     xlab = "Jumlah Cluster (Jumlah Cluster Paling Optimal = 2)",
     ylab = "Average Silhouette Score",
     main = "Silhouette Method for Optimal k")
```

Silhouette Method for Optimal k



K-Means

```
set.seed(123)
kmeans_result= kmeans(
  clean_data_scaled,
  centers=2,
  nstart=25
)

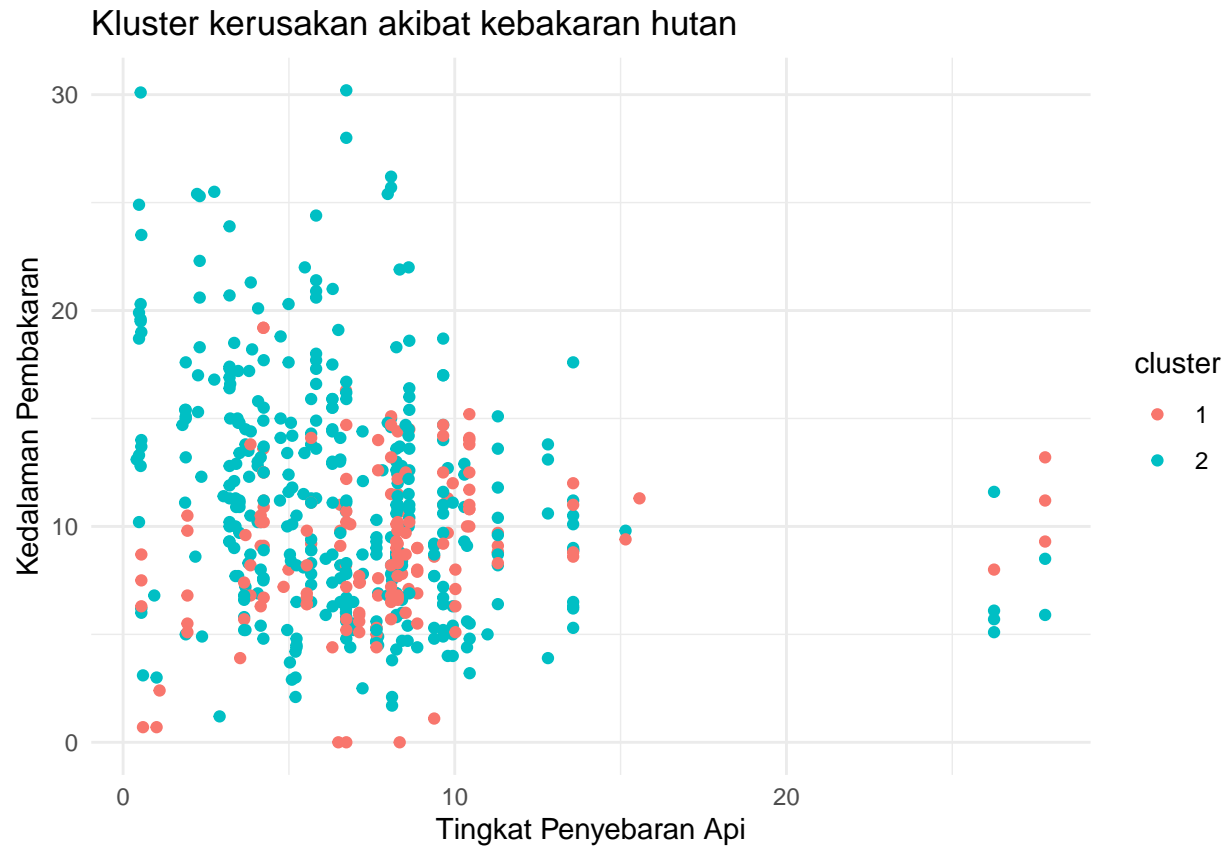
clean_data$cluster= as.factor(kmeans_result$cluster)
tail(clean_data)
```

```
##      kedalaman_pembakaran proporsi_lapisan_organik_terbakar karbon_terbakar
## 1156          12.6              0.239          4252.5
## 1167          22.0              0.499          7085.8
## 1168          14.5              0.644          4609.7
## 1169          14.2              0.441          4692.3
## 1170          11.5              0.696          4875.3
## 1171          12.2              0.709          3774.1
##      biomassa_terbakar karbon_residu_tanah kedalaman_rata_rata_lapisan_residu
## 1156          1159.455          15007.5              40
## 1167           332.823           8317.7              22
## 1168           616.318           2797.3              8
```

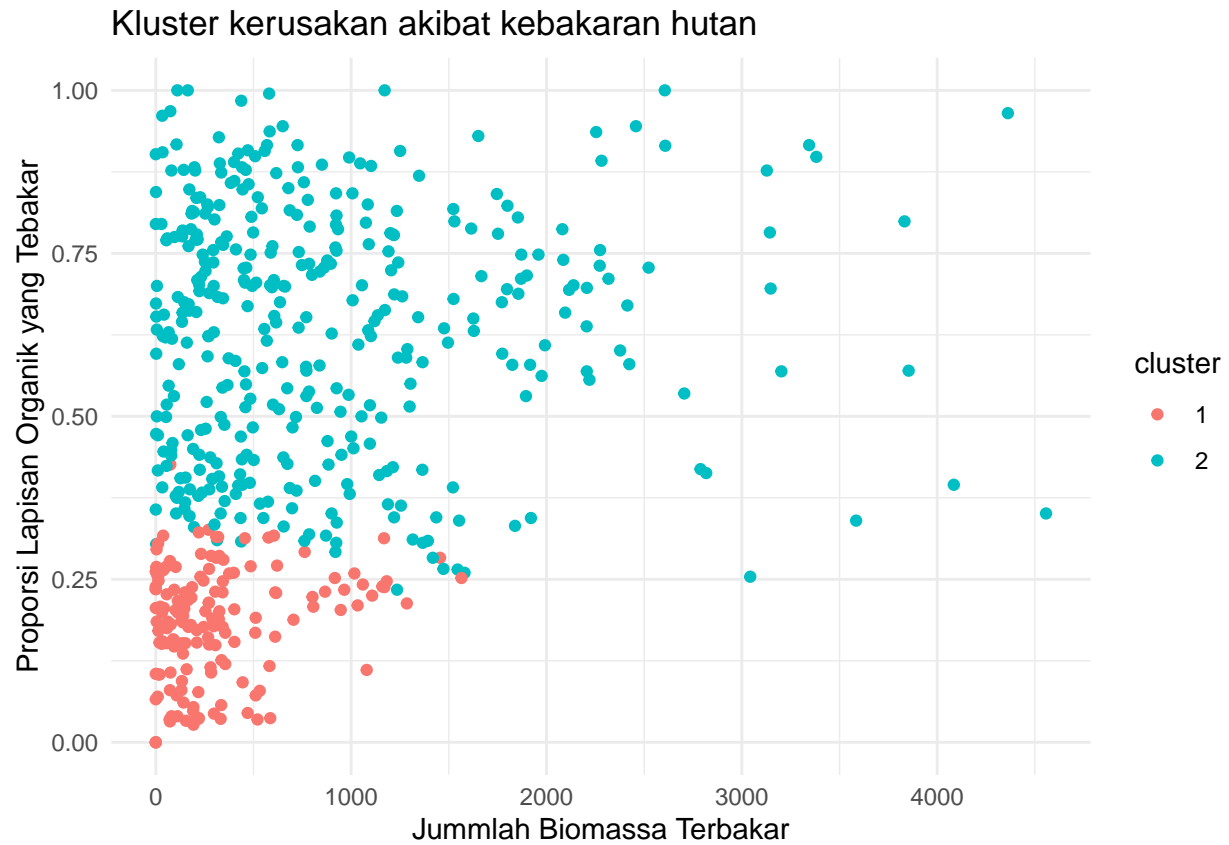
## 1169	950.460	6480.6	18
## 1170	3147.472	1683.1	5
## 1171	451.241	1695.4	5
##	tingkat_penyebaran_api	cluster	
## 1156	7.69	1	
## 1167	8.61	2	
## 1168	8.61	2	
## 1169	8.61	2	
## 1170	8.61	2	
## 1171	8.61	2	

Visualisasi Data

```
ggplot(
  clean_data,
  aes(
    x = tingkat_penyebaran_api,
    y = kedalaman_pembakaran,
    color = cluster
  )
) + geom_point() + labs(
  title = "Kluster kerusakan akibat kebakaran hutan",
  x="Tingkat Penyebaran Api",
  y="Kedalaman Pembakaran"
) + theme_minimal()
```

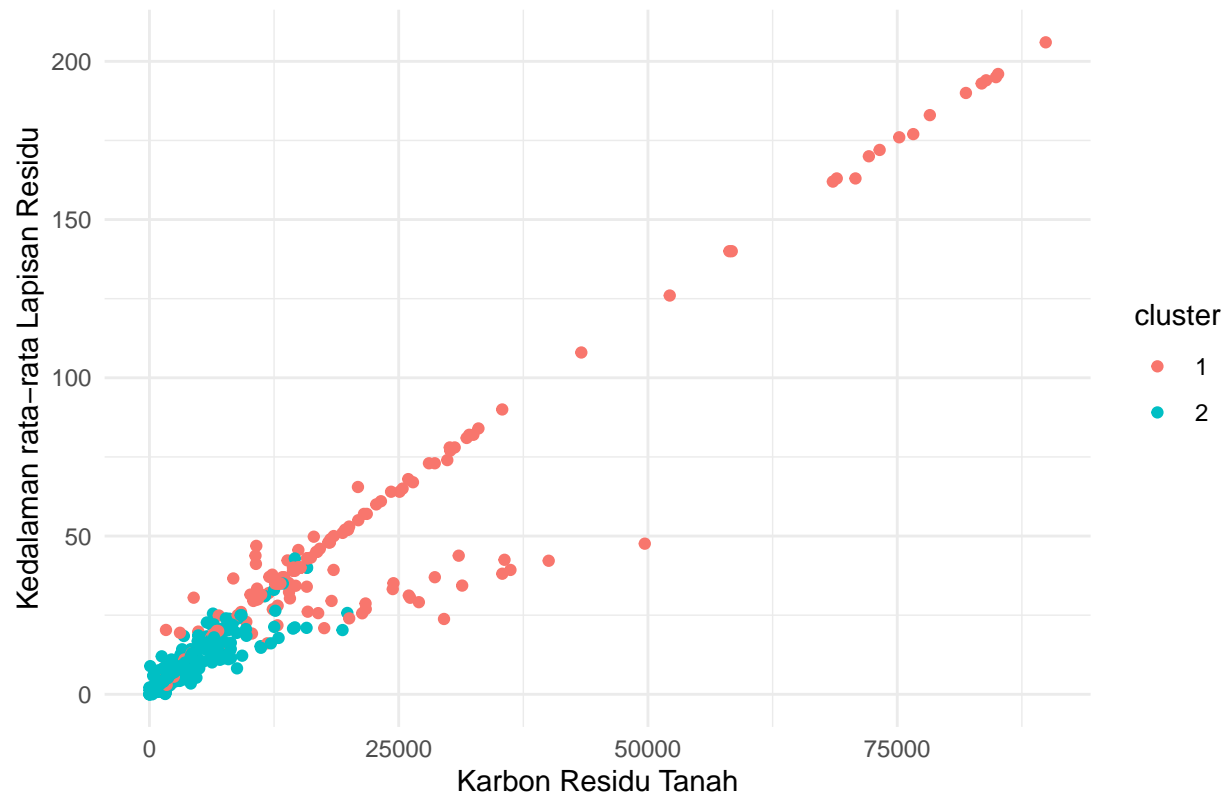


```
ggplot(  
  clean_data,  
  aes(  
    x = biomassa_terbakar,  
    y = proporsi_lapisan_organik_terbakar,  
    color = cluster  
  )  
)+ geom_point() + labs(  
  title = "Kluster kerusakan akibat kebakaran hutan",  
  x="Jumlah Biomassa Terbakar",  
  y="Proporsi Lapisan Organik yang Terbakar"  
)+ theme_minimal()
```

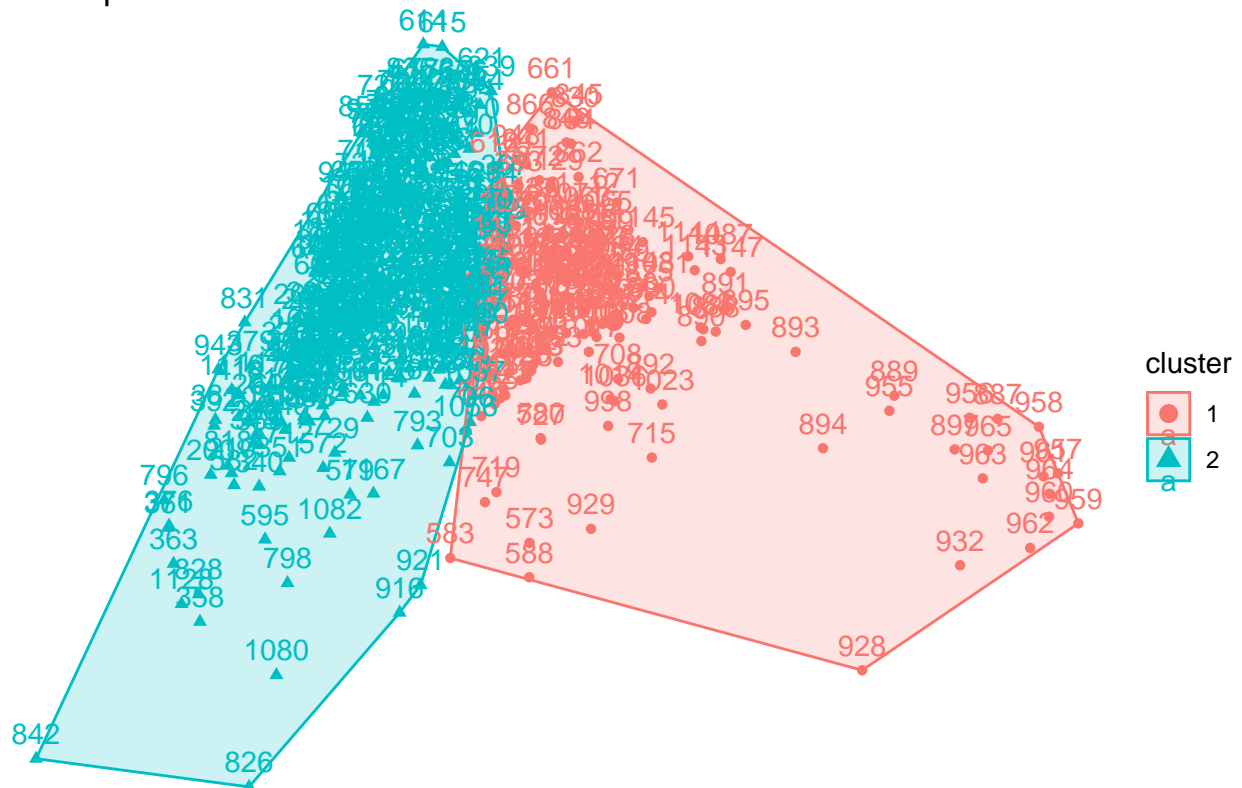
```
ggplot(
  clean_data,
  aes(
    x = karbon_residu_tanah,
    y = kedalaman_rata_rata_lapisan_residu,
    color = cluster
  )
) + geom_point() + labs(
  title = "Kluster kerusakan akibat kebakaran hutan",
  x="Karbon Residu Tanah",
  y="Kedalaman rata-rata Lapisan Residu"
) + theme_minimal()
```

Kluster kerusakan akibat kebakaran hutan



```
fviz_cluster(kmeans_result, data = clean_data_scaled) + theme_void()
```

Cluster plot



```
cluster_summary = clean_data %>%
  select(kedalaman_pembakaran, proporsi_lapisan_organik_terbakar,
         karbon_terbakar, biomassa_terbakar, karbon_residu_tanah,
         kedalaman_rata_rata_lapisan_residu, tingkat_penyebaran_api, cluster) %>%
  group_by(cluster) %>%
  summarise(
    kedalaman_pembakaran = mean(kedalaman_pembakaran),
    karbon_residu_tanah = mean(karbon_residu_tanah),
    kedalaman_lapisan_residu = mean(kedalaman_rata_rata_lapisan_residu),
    tingkat_penyebaran_api = mean(tingkat_penyebaran_api),
    proporsi_lapisan_organik_terbakar = mean(proporsi_lapisan_organik_terbakar),
    karbon_terbakar = mean(karbon_terbakar),
    biomassa_terbakar = mean(biomassa_terbakar)
  )

cluster_summary
```

```
## # A tibble: 2 x 8
##   cluster kedalaman_pembakaran karbon_residu_tanah kedalaman_lapisan_residu
##   <fct>          <dbl>          <dbl>          <dbl>
## 1 1              9.01          22830.          53.3
## 2 2             11.3           3261.           8.18
## # i 4 more variables: tingkat_penyebaran_api <dbl>,
## #   proporsi_lapisan_organik_terbakar <dbl>, karbon_terbakar <dbl>,
## #   biomassa_terbakar <dbl>
```