

# Modul\_Analisis\_Kuantitatif

Roni Yunis

2023-10-12

## 1. Deskripsi

Tujuan analisis adalah menghasilkan model prediksi tanaman pangan melalui analisis klustering. Temuan dari analisis ini akan menjadi landasan penting dalam pengambilan keputusan untuk meningkatkan ketahanan pangan khususnya di wilayah Sumatera Utara. Untuk pengujian dan analisis akan menggunakan data tanaman pangan padi dari 2010-2022.

## 2. Load Library

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'  
  
## The following objects are masked from 'package:base':  
##  
##   date, intersect, setdiff, union
```

```
library(summarytools)  
library(readxl)  
library(cluster)  
library(scales)  
library(factoextra)
```

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

```
library(mclust)
```

```
## Package 'mclust' version 6.0.0  
## Type 'citation("mclust")' for citing this R package in publications.
```

```
#library(vars)  
#library(forecast)  
#library(caret)  
#library(timetk)
```

### 3. Analisis Klustering Produksi Padi

#### Obtain - Load Dataset

```
padi_sumut <- read_excel("data/bps_padi_sumut.xlsx")  
padi_sumut
```

```
## # A tibble: 442 x 5  
##   Tahun 'Kabupaten Kota' 'Rata-rata produksi' Produksi 'Luas Panen'  
##   <dbl> <chr>           <chr>           <chr>           <chr>  
## 1 2022 Asahan          61,64          62786.65        10185.41  
## 2 2022 Batu Bara       55,4          71050.570000000007 12827.29  
## 3 2022 Binjai          50,34          6266.34         1244.910000~  
## 4 2022 Dairi           49,2          38714.36         7868.1  
## 5 2022 Deli Serdang    60,91          328854.78999999998 53984.69  
## 6 2022 Gunungsitoli    51,6          11017.47         2135.570000~  
## 7 2022 Humbang Hasundutan 41,95          75462.0800000000002 17992.27  
## 8 2022 Karo            70,23          69058.42         9834.459999~  
## 9 2022 Labuanbatu Utara  40,37          80203.7          19868.18999~  
## 10 2022 Labuhan Batu    38,99          83640.899999999994 21455.81  
## # i 432 more rows
```

#### Scrub Data

Proses Scrub yang akan dilakukan adalah merubah type data dan menghapus data yang tidak penting atau data yang kosng (not available). Berdasarkan data\_sumut dapat dilihat bahwa type data dari kolom Tahun adalah double, rata-rata produksi, produksi dan luas panen bertipe karakter. Maka untuk mendukung proses analisis maka type data Tahun harus dirubah ke type Date, dan data yg bertipe karakter dirubah kedalam type numeric

##### a. Merubah type data

```
# Merubah type data karakter menjadi numeric  
padi_sumut$`Rata-rata produksi` <- as.numeric(padi_sumut$`Rata-rata produksi`)
```

```
## Warning: NAs introduced by coercion
```

```
padi_sumut$Produksi <- as.numeric(padi_sumut$Produksi)
```

```
## Warning: NAs introduced by coercion
```

```
padi_sumut$`Luas Panen` <- as.numeric(padi_sumut$`Luas Panen`)
```

```
## Warning: NAs introduced by coercion
```

```
padi_sumut
```

```
## # A tibble: 442 x 5
##   Tahun 'Kabupaten Kota' 'Rata-rata produksi' Produksi 'Luas Panen'
##   <dbl> <chr>              <dbl>      <dbl>      <dbl>
## 1 2022 Asahan              NA    62787.    10185.
## 2 2022 Batu Bara          NA    71051.    12827.
## 3 2022 Binjai             NA     6266.     1245.
## 4 2022 Dairi              NA    38714.     7868.
## 5 2022 Deli Serdang       NA   328855.    53985.
## 6 2022 Gunungsitoli       NA    11017.     2136.
## 7 2022 Humbang Hasundutan NA    75462.    17992.
## 8 2022 Karo              NA    69058.     9834.
## 9 2022 Labuanbatu Utara    NA    80204.    19868.
## 10 2022 Labuhan Batu      NA    83641.    21456.
## # i 432 more rows
```

```
# Merubah type Tahun menjadi Date
```

```
padi_sumut$Tahun <- make_date(padi_sumut$Tahun)
glimpse(padi_sumut)
```

```
## Rows: 442
## Columns: 5
## $ Tahun          <date> 2022-01-01, 2022-01-01, 2022-01-01, 2022-01-01, ~
## $ 'Kabupaten Kota' <chr> "Asahan", "Batu Bara", "Binjai", "Dairi", "Deli S~
## $ 'Rata-rata produksi' <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Produksi        <dbl> 62786.65, 71050.57, 6266.34, 38714.36, 328854.79, ~
## $ 'Luas Panen'     <dbl> 10185.41, 12827.29, 1244.91, 7868.10, 53984.69, 2~
```

```
# Merubah nama variabel Kabupaten Kota, Rata-rata Produksi dan Luas Panen
```

```
names(padi_sumut)[names(padi_sumut) == "Kabupaten Kota"] <- "Kabupaten_kota"
names(padi_sumut)[names(padi_sumut) == "Rata-rata produksi"] <- "Rata_rata_produksi"
names(padi_sumut)[names(padi_sumut) == "Luas Panen"] <- "Luas_panen"
glimpse(padi_sumut)
```

```
## Rows: 442
## Columns: 5
## $ Tahun          <date> 2022-01-01, 2022-01-01, 2022-01-01, 2022-01-01, 20~
## $ Kabupaten_kota <chr> "Asahan", "Batu Bara", "Binjai", "Dairi", "Deli Ser~
## $ Rata_rata_produksi <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Produksi        <dbl> 62786.65, 71050.57, 6266.34, 38714.36, 328854.79, 1~
## $ Luas_panen      <dbl> 10185.41, 12827.29, 1244.91, 7868.10, 53984.69, 213~
```

```
tail(padi_sumut)
```

```
## # A tibble: 6 x 5
##   Tahun      Kabupaten_kota Rata_rata_produksi Produksi Luas_panen
##   <date>      <chr>              <dbl>      <dbl>      <dbl>
## 1 2010-01-01 Tanjungbalai          45.5        1942        427
## 2 2010-01-01 Tapanuli Selatan      48.8       143348       29398
## 3 2010-01-01 Tapanuli Tengah       43.4       130389       30039
## 4 2010-01-01 Tapanuli Utara         44.3       119723       27030
## 5 2010-01-01 Tebing Tinggi         48.2         5474        1136
## 6 2010-01-01 Toba Samosir          47.4       106075       22353
```

```
head(padi_sumut)
```

```
## # A tibble: 6 x 5
##   Tahun      Kabupaten_kota Rata_rata_produksi Produksi Luas_panen
##   <date>      <chr>              <dbl>      <dbl>      <dbl>
## 1 2022-01-01 Asahan              NA       62787.       10185.
## 2 2022-01-01 Batu Bara           NA       71051.       12827.
## 3 2022-01-01 Binjai              NA        6266.        1245.
## 4 2022-01-01 Dairi              NA       38714.        7868.
## 5 2022-01-01 Deli Serdang        NA      328855.       53985.
## 6 2022-01-01 Gunungsitoli        NA       11017.        2136.
```

## b. Menghapus Data Kosong

```
# Menampilkan variabel dengan baris kosong
colSums(is.na(padi_sumut))
```

```
##           Tahun      Kabupaten_kota Rata_rata_produksi      Produksi
##           0              0              67              8
##   Luas_panen
##           9
```

```
# Menghapus data NA's
padisumut_clean <- na.omit(padi_sumut)
summary(padisumut_clean)
```

```
##           Tahun      Kabupaten_kota      Rata_rata_produksi      Produksi
##   Min.   :2010-01-01 Length:374      Min.   : 0.00      Min.   :      0
##   1st Qu.:2012-01-01 Class :character 1st Qu.:42.26      1st Qu.: 19093
##   Median :2015-01-01 Mode  :character Median :47.89      Median : 68683
##   Mean   :2015-01-15      Mean   :46.83      Mean   : 211354
##   3rd Qu.:2018-01-01      3rd Qu.:52.29      3rd Qu.: 147796
##   Max.   :2022-01-01      Max.   :71.00      Max.   :5136186
##   Luas_panen
##   Min.   :      0
##   1st Qu.: 3895
##   Median : 15568
```

```
## Mean : 41914
## 3rd Qu.: 29879
## Max. :988068
```

```
colSums(is.na(padisumut_clean))
```

```
##          Tahun      Kabupaten_kota Rata_rata_produksi      Produksi
##          0          0          0          0
##      Luas_panen
##          0
```

Semua data NA's sudah dihapus. Data padi di wilayah sumut mulai dari Tahun 2010 - 2022 dengan frekuensi data per Tahun. Dalam dataset padisumut\_clean masih terdapat data dari provinsi Sumatera Utara keseluruhan. Agar memudahkan dalam tahap analisis selanjutnya maka data padisumut\_clean yang diambil hanya berdasarkan kabupaten/kota saja. Sekarang data final yang siap dianalisis adalah sebanyak 374 baris observasi dengan 5 kolom variabel

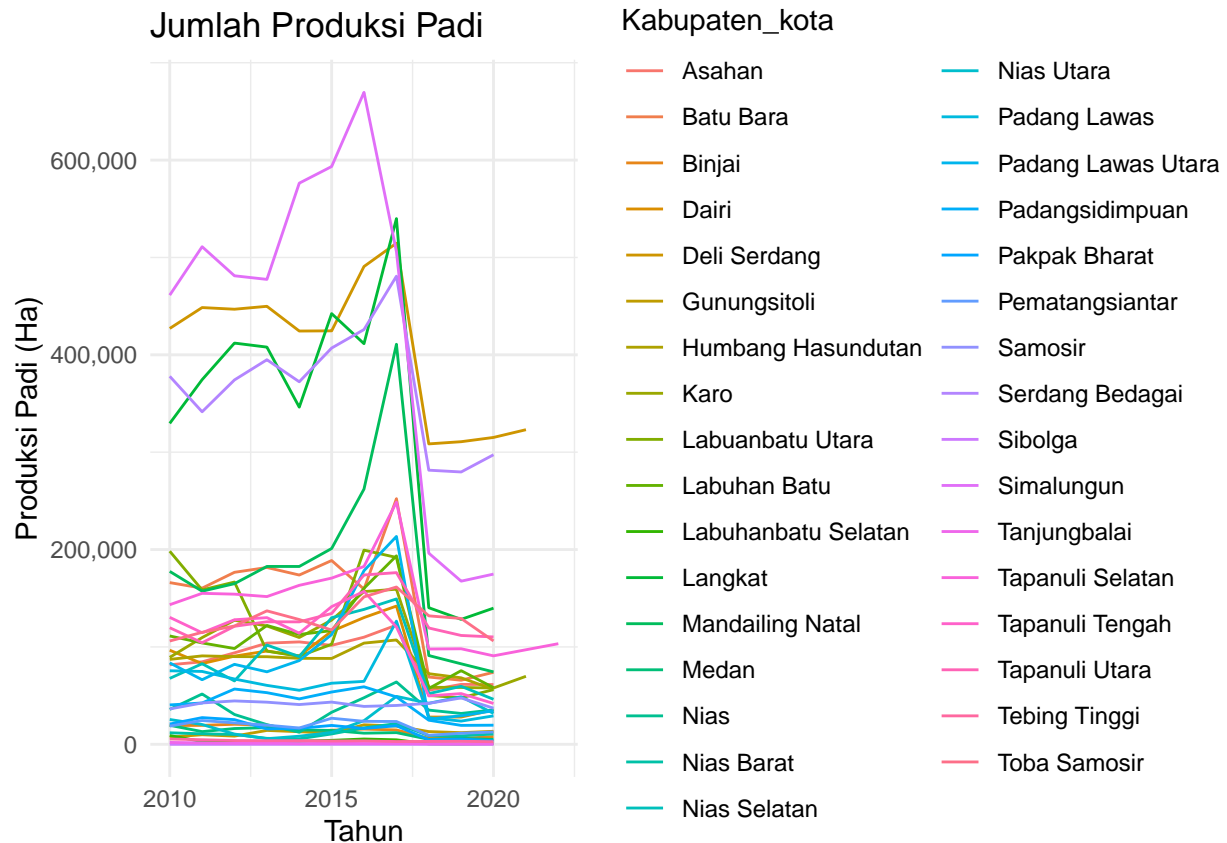
```
padisumutfilter <- padisumut_clean[padisumut_clean$Kabupaten_kota != "Sumatera Utara",]
padisumutfilter
```

```
## # A tibble: 362 x 5
##   Tahun      Kabupaten_kota      Rata_rata_produksi Produksi Luas_panen
##   <date>      <chr>              <dbl>      <dbl>      <dbl>
## 1 2022-01-01 Tapanuli Selatan      50      103327.    20802.
## 2 2021-01-01 Deli Serdang        60      323108     53981
## 3 2021-01-01 Karo                71       69829      9844
## 4 2020-01-01 Asahan              57.1     61350.    10737.
## 5 2020-01-01 Batu Bara           56.9     73939.    12988.
## 6 2020-01-01 Binjai              54.0      7870.     1456.
## 7 2020-01-01 Dairi               53.9     35311.     6546.
## 8 2020-01-01 Deli Serdang        63.5    315156.    49658.
## 9 2020-01-01 Gunungsitoli        56.8     13352.     2349.
## 10 2020-01-01 Humbang Hasundutan  47.1     56390.    11969.
## # i 352 more rows
```

## Eksplorasi Data Analisis (EDA)

### Jumlah Produksi Padi per Kabupaten/Kota

```
plot_1 <- ggplot(padisumutfilter, aes(x = Tahun, y = Produksi, group = Kabupaten_kota, color = Kabupaten_kota)) +
  geom_line() +
  labs(title = "Jumlah Produksi Padi", x = "Tahun", y = "Produksi Padi (Ha)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
plot_1
```



```
# Hitung total produksi per kabupaten/kota
totproduksi <- padisumutfilter %>%
  mutate(first_date_month = floor_date(Tahun, unit = "year")) %>%
  group_by(Kabupaten_kota) %>%
  summarise(jumlahproduksi = sum(Produksi)) %>%
  arrange(jumlahproduksi)
totproduksi
```

```
## # A tibble: 33 x 2
##   Kabupaten_kota      jumlahproduksi
##   <chr>              <dbl>
## 1 Sibolga              0
## 2 Tanjungbalai        10392.
## 3 Labuhanbatu Selatan  36415.
## 4 Tebing Tinggi       40051.
## 5 Nias Barat          121294.
## 6 Medan               134451.
## 7 Gunungsitoli        141274.
## 8 Binjai              156803.
## 9 Pakpak Bharat       174220.
## 10 Pematangsiantar    208503.
## # i 23 more rows
```

```
# Turn off dplyr summarise() warning
options(dplyr.summarise.inform = FALSE)
```

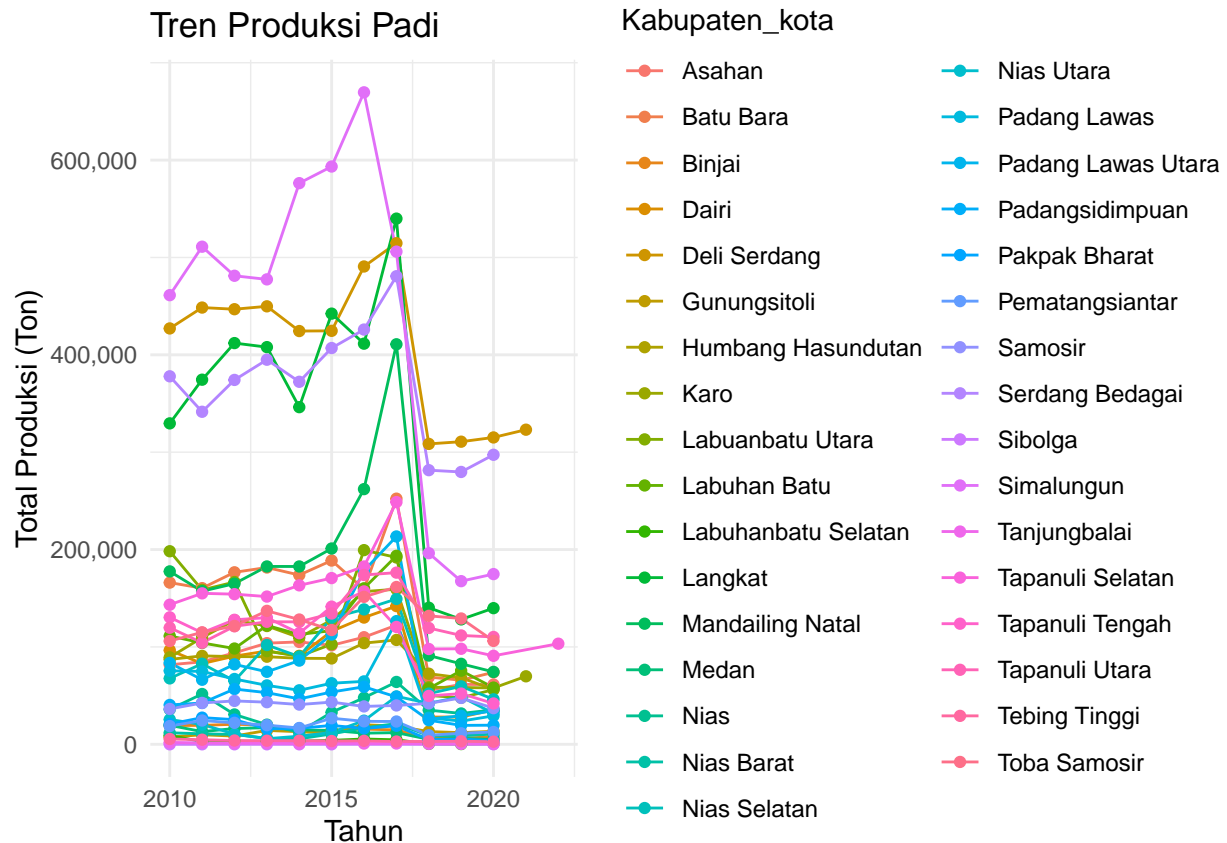
```
# Hitung total produksi per kabupaten/tahun
totproduksi <- padisumutfilter %>%
  group_by(Tahun, Kabupaten_kota, .drop = FALSE) %>% # Add .drop argument
  summarise(total_produksi = sum(Produksi))
```

```
totproduksi
```

```
## # A tibble: 362 x 3
## # Groups:   Tahun [13]
##   Tahun      Kabupaten_kota    total_produksi
##   <date>      <chr>              <dbl>
## 1 2010-01-01 Asahan                81685
## 2 2010-01-01 Batu Bara            166063
## 3 2010-01-01 Binjai                19247
## 4 2010-01-01 Dairi                 96612
## 5 2010-01-01 Deli Serdang         427104
## 6 2010-01-01 Gunungsitoli          7387
## 7 2010-01-01 Humbang Hasundutan    87205
## 8 2010-01-01 Karo                  89302
## 9 2010-01-01 Labuanbatu Utara       198284
## 10 2010-01-01 Labuhan Batu         111260
## # i 352 more rows
```

### Tren Produksi Padi per Kabupaten/Kota

```
# Visualisasi Tren Produksi Padi
plot_2 <- ggplot(totproduksi, aes(x = Tahun, y = total_produksi, color = Kabupaten_kota)) +
  geom_line() +
  geom_point() +
  labs(title = "Tren Produksi Padi", x = "Tahun", y = "Total Produksi (Ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
plot_2
```



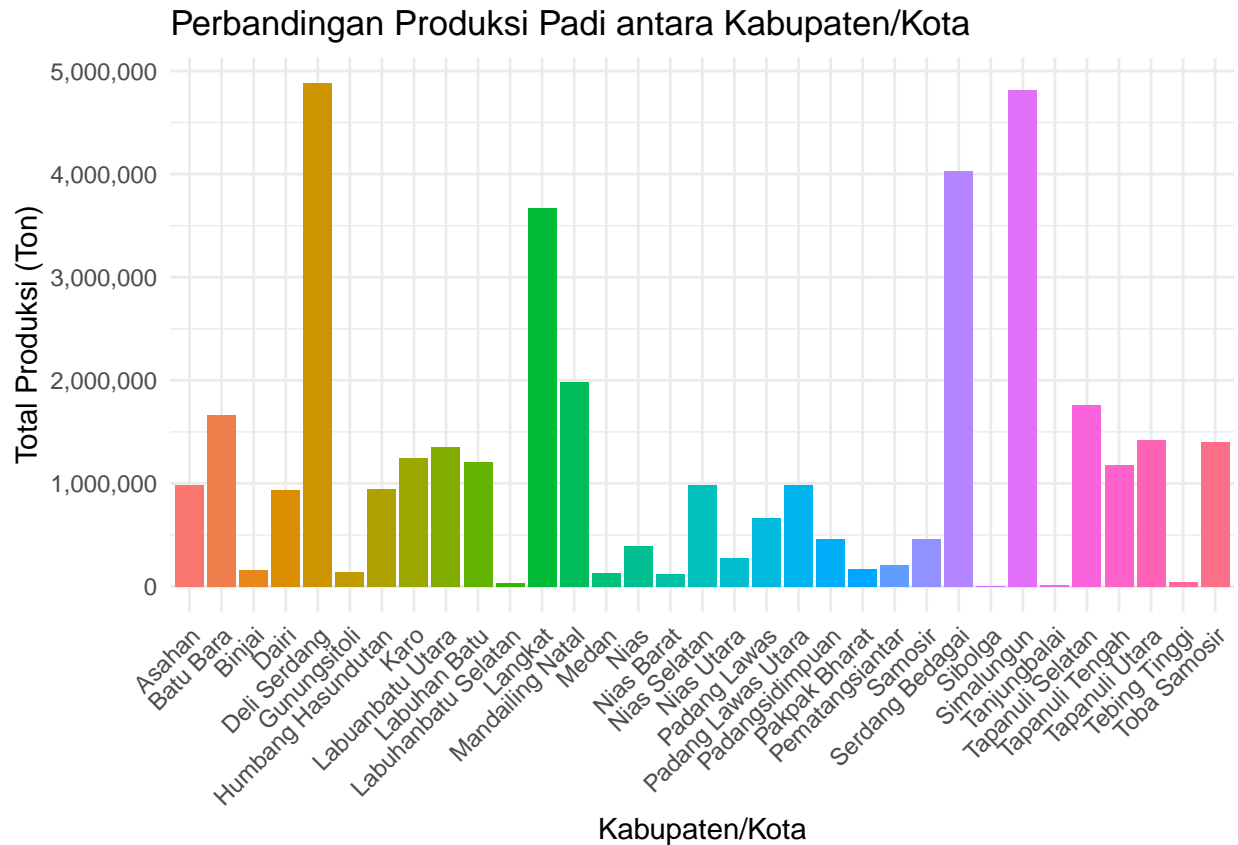
### Perbandingan Produksi Padi per Kabupaten/Kota

```
# Hitung total produksi padi per Kabupaten/Kota
totproduksi_kabkota <- padisumutfilter %>%
  group_by(Kabupaten_kota) %>%
  summarise(total_produksi = sum(Produksi))

# Visualisasi dengan barplot
plot_3 <- ggplot(totproduksi_kabkota, aes(x = Kabupaten_kota, y = total_produksi, fill = Kabupaten_kota)) +
  geom_bar(stat = "identity") +
  labs(title = "Perbandingan Produksi Padi antara Kabupaten/Kota", x = "Kabupaten/Kota", y = "Total Produksi") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "none") # hilangkan legend

# Rotasi sumbu x agar mudah dibaca
plot_3 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```





Perbandingan produksi padi dalam kurun waktu 2010 - 2022 adalah dari Kabupaten Deli Serdang yang terbanyak dengan Jumlah Produksi Padi sebesar 4884323.62 Ton

### Perbandingan Rata-rata produksi Padi per Kabupaten/Kota

```
# Turn off dplyr summarise() warning
options(dplyr.summarise.inform = FALSE)

# Melihat tren rata-rata produksi padi
rataproduksi <- padisumutfilter %>%
  group_by(Tahun, Kabupaten_kota, .drop = FALSE) %>% # Add .drop argument
  summarise(Rata_rata_produksi = sum(Rata_rata_produksi))
rataproduksi
```

```
## # A tibble: 362 x 3
## # Groups:   Tahun [13]
##   Tahun      Kabupaten_kota      Rata_rata_produksi
##   <date>      <chr>              <dbl>
## 1 2010-01-01 Asahan                47.8
## 2 2010-01-01 Batu Bara              48.5
## 3 2010-01-01 Binjai                 47.7
## 4 2010-01-01 Dairi                  41.0
## 5 2010-01-01 Deli Serdang           50.3
## 6 2010-01-01 Gunungsitoli           40.7
```

```
## 7 2010-01-01 Humbang Hasundutan 45.2
## 8 2010-01-01 Karo 38.5
## 9 2010-01-01 Labuanbatu Utara 48.2
## 10 2010-01-01 Labuhan Batu 48.2
## # i 352 more rows
```

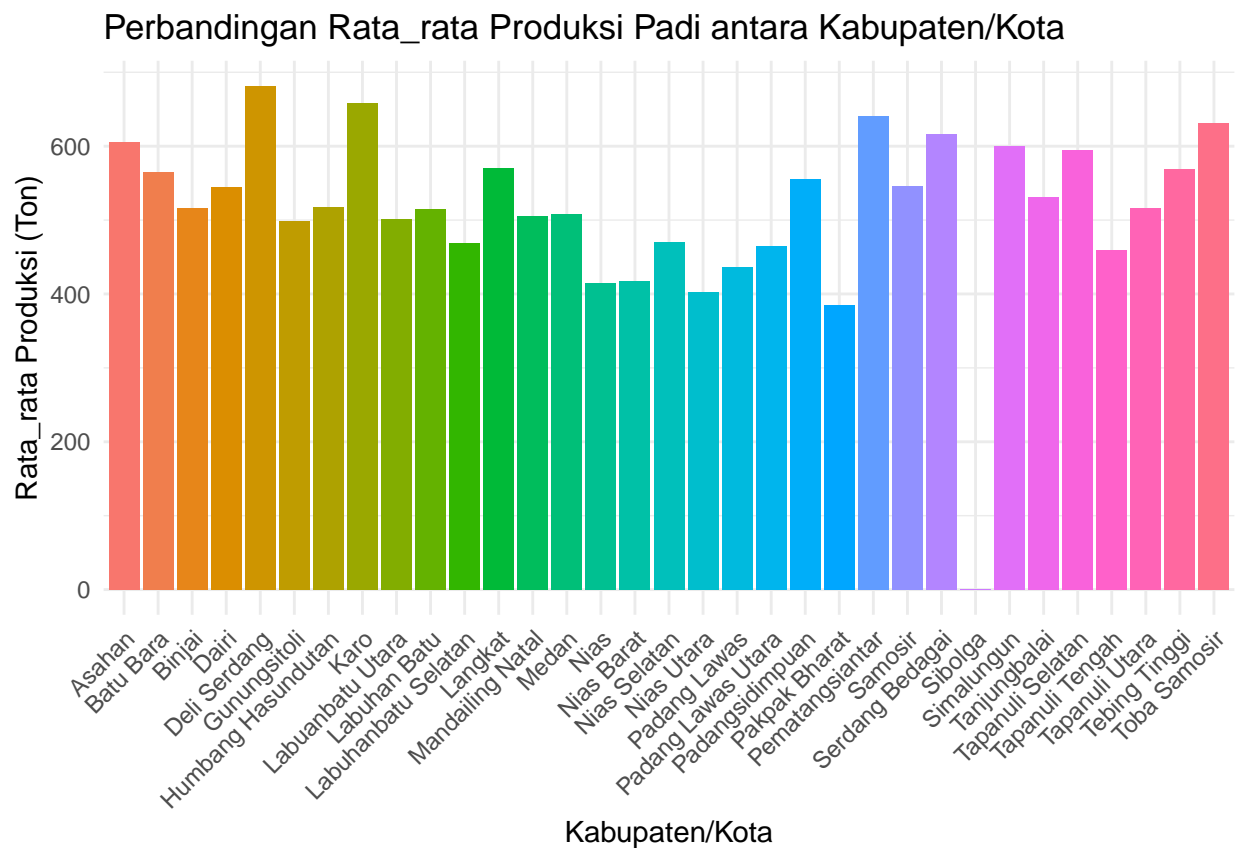
```
# Hitung total rata-rata produksi padi per Kabupaten/Kota
totrataproduksi_kabkota <- padisumutfilter %>%
  group_by(Kabupaten_kota) %>%
  summarise(total_rataproduksi = sum(Rata_rata_produksi))
```

```
# Visualisasi dengan barplot
```

```
plot_4 <- ggplot(totrataproduksi_kabkota, aes(x = Kabupaten_kota, y = total_rataproduksi, fill = Kabupaten_kota)) +
  geom_bar(stat = "identity") +
  labs(title = "Perbandingan Rata_rata Produksi Padi antara Kabupaten/Kota", x = "Kabupaten/Kota", y = "Rata_rata Produksi (Ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "none") # hilangkan legend
```

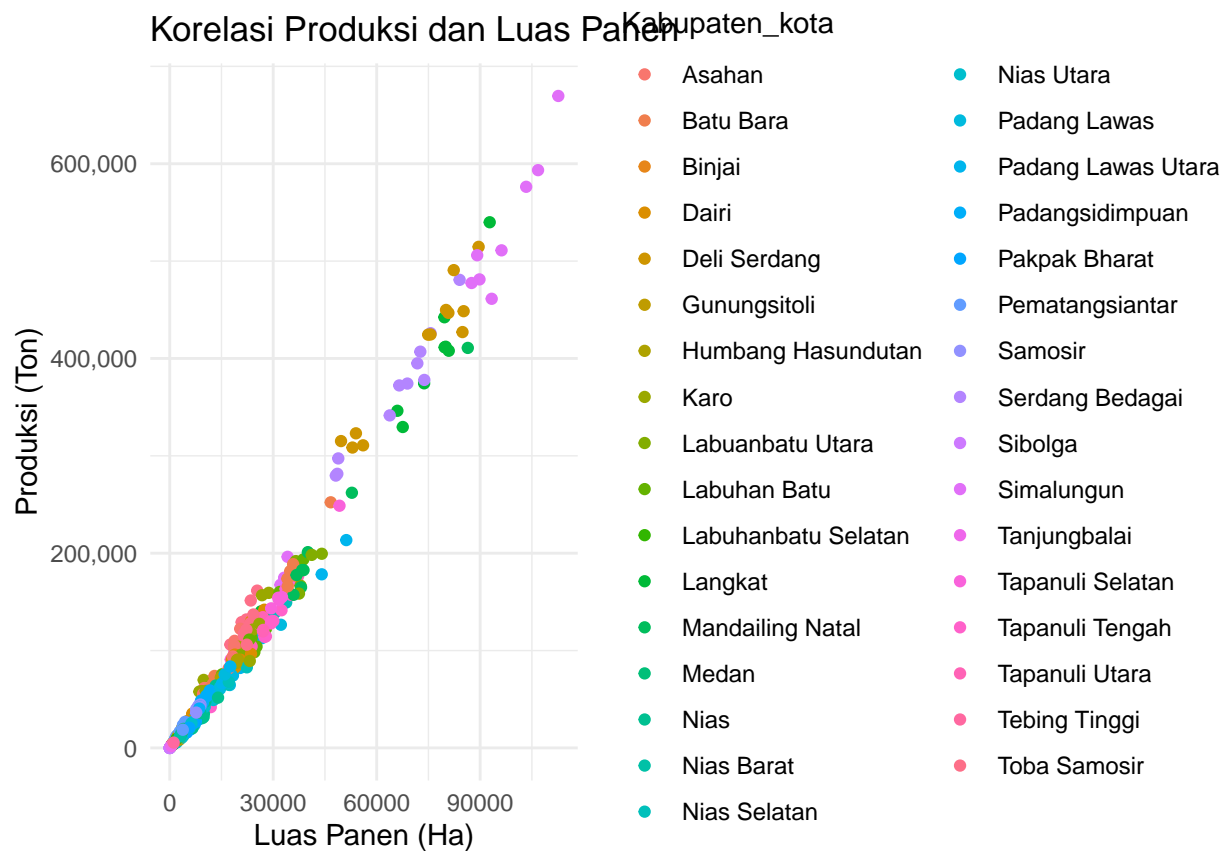
```
# Rotasi sumbu x agar mudah dibaca
```

```
plot_4 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## Korelasi Luas Panen dan Produksi Padi

```
# Visualisasi korelasi luas panen dan produksi padi
plot_5 <- ggplot(padisumutfilter, aes(x = Luas_panen, y = Produksi, color = Kabupaten_kota)) +
  geom_point() +
  labs(title = "Korelasi Produksi dan Luas Panen", x = "Luas Panen (Ha)", y = "Produksi (Ton)") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "right")
plot_5
```



```
# Hitung Pearson correlation coefficient
correlation_padi <- cor(padisumutfilter$Luas_panen, padisumutfilter$Produksi)

# Nilai correlation coefficient
cat("Koefisien Korelasi Pearson Luas Panen dan Produksi:", correlation_padi)
```

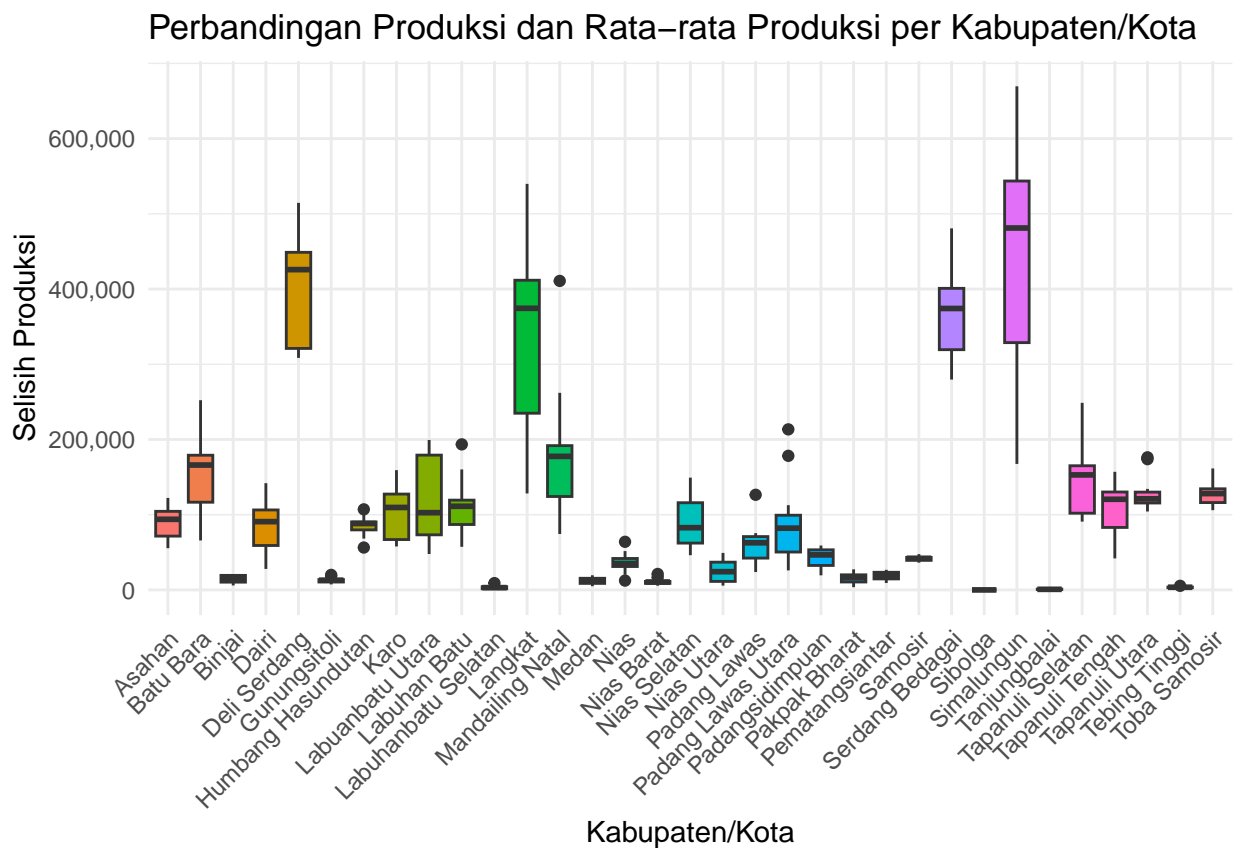
```
## Koefisien Korelasi Pearson Luas Panen dan Produksi: 0.9928962
```

Berdasarkan grafik diatas bisa lihat bahwa ada korelasi antara luas panen dan produksi. Hubungan korelasi adalah positif dengan nilai korelasi sebesar 0,99, sehingga bisa disimpulkan bahwa jika semakin banyak luas panen maka akan meningkat jumlah produksi padi.

## Perbandingan Rata-rata produksi dengan Produksi Padi

```
# Create a boxplot using ggplot2
plot_6 <- ggplot(padisumutfilter, aes(x = Kabupaten_kota, y = Produksi - Rata_rata_produksi, fill = Kabupaten_kota)) +
  geom_boxplot() +
  labs(title = "Perbandingan Produksi dan Rata-rata Produksi per Kabupaten/Kota", x = "Kabupaten/Kota",
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(legend.position = "none")

# Rotasi sumbu x agar mudah dibaca
plot_6 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

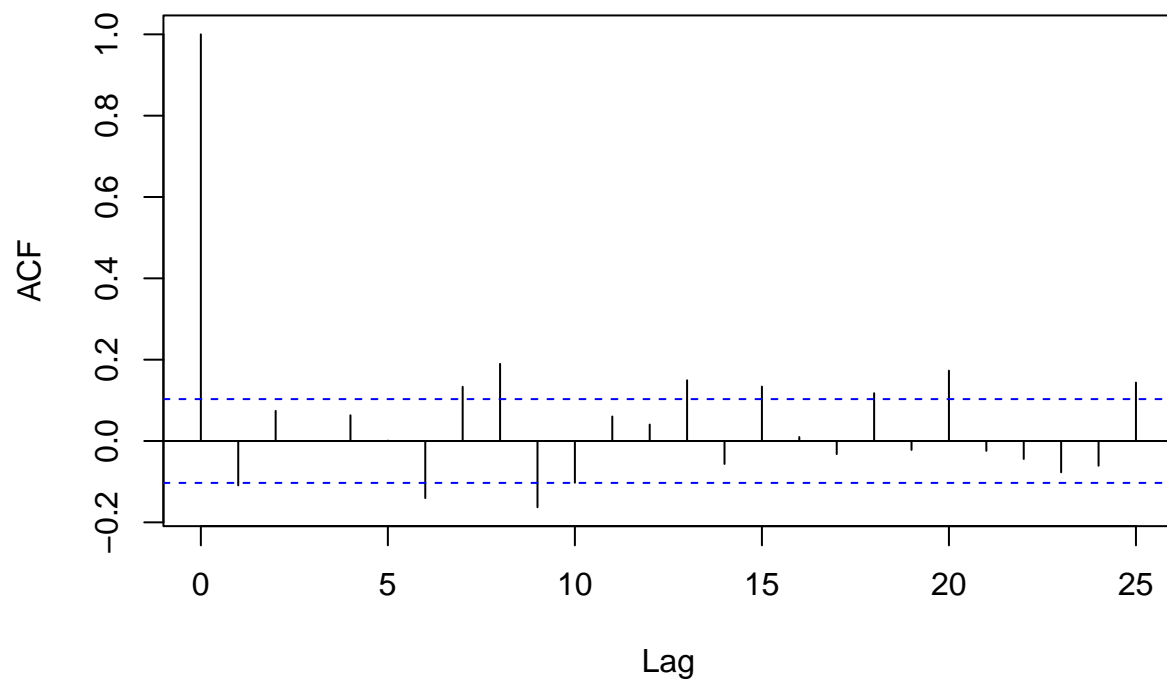


Melihat pola musiman dalam data untuk produksi padi

```
# Convert Tahun to Date format
padisumutfilter$Tahun <- as.Date(paste0(padisumutfilter$Tahun, "-01-01"))

# Create ACF plot
acf_plot <- acf(padisumutfilter$Produksi, main = "ACF Plot of Produksi Padi")
```

## ACF Plot of Produksi Padi

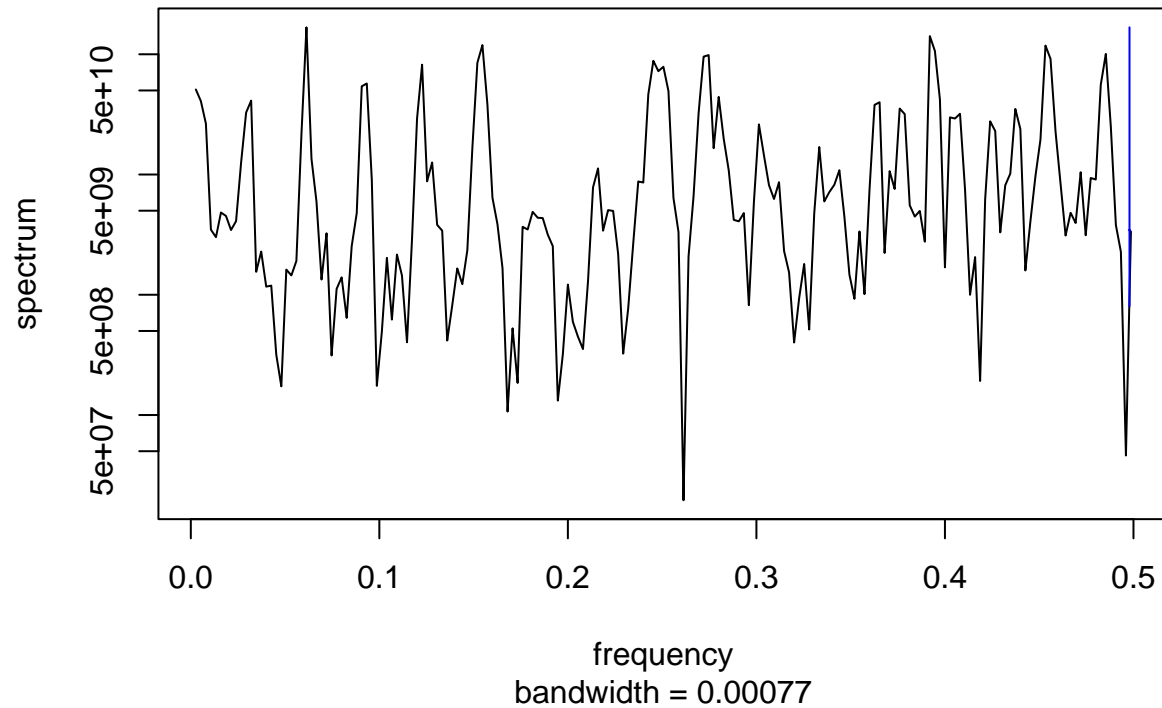


```
# Print ACF plot  
print(acf_plot)
```

```
##  
## Autocorrelations of series 'padisumutfilter$Produksi', by lag  
##  
##      0      1      2      3      4      5      6      7      8      9     10  
## 1.000 -0.109  0.074  0.000  0.063  0.001 -0.140  0.133  0.190 -0.163 -0.103  
##     11     12     13     14     15     16     17     18     19     20     21  
## 0.060  0.040  0.149 -0.056  0.134  0.010 -0.032  0.118 -0.022  0.173 -0.024  
##     22     23     24     25  
## -0.044 -0.077 -0.061  0.144
```

```
# Create periodogram  
periodogram_plot <- spectrum(padisumutfilter$Produksi, main = "Periodogram of Produksi Padi")
```

## Periodogram of Produksi Padi



```
# Print periodogram
periodogram_plot
```

```
## $freq
## [1] 0.002666667 0.005333333 0.008000000 0.010666667 0.013333333 0.016000000
## [7] 0.018666667 0.021333333 0.024000000 0.026666667 0.029333333 0.032000000
## [13] 0.034666667 0.037333333 0.040000000 0.042666667 0.045333333 0.048000000
## [19] 0.050666667 0.053333333 0.056000000 0.058666667 0.061333333 0.064000000
## [25] 0.066666667 0.069333333 0.072000000 0.074666667 0.077333333 0.080000000
## [31] 0.082666667 0.085333333 0.088000000 0.090666667 0.093333333 0.096000000
## [37] 0.098666667 0.101333333 0.104000000 0.106666667 0.109333333 0.112000000
## [43] 0.114666667 0.117333333 0.120000000 0.122666667 0.125333333 0.128000000
## [49] 0.130666667 0.133333333 0.136000000 0.138666667 0.141333333 0.144000000
## [55] 0.146666667 0.149333333 0.152000000 0.154666667 0.157333333 0.160000000
## [61] 0.162666667 0.165333333 0.168000000 0.170666667 0.173333333 0.176000000
## [67] 0.178666667 0.181333333 0.184000000 0.186666667 0.189333333 0.192000000
## [73] 0.194666667 0.197333333 0.200000000 0.202666667 0.205333333 0.208000000
## [79] 0.210666667 0.213333333 0.216000000 0.218666667 0.221333333 0.224000000
## [85] 0.226666667 0.229333333 0.232000000 0.234666667 0.237333333 0.240000000
## [91] 0.242666667 0.245333333 0.248000000 0.250666667 0.253333333 0.256000000
## [97] 0.258666667 0.261333333 0.264000000 0.266666667 0.269333333 0.272000000
## [103] 0.274666667 0.277333333 0.280000000 0.282666667 0.285333333 0.288000000
## [109] 0.290666667 0.293333333 0.296000000 0.298666667 0.301333333 0.304000000
## [115] 0.306666667 0.309333333 0.312000000 0.314666667 0.317333333 0.320000000
## [121] 0.322666667 0.325333333 0.328000000 0.330666667 0.333333333 0.336000000
## [127] 0.338666667 0.341333333 0.344000000 0.346666667 0.349333333 0.352000000
```

```

## [133] 0.354666667 0.357333333 0.360000000 0.362666667 0.365333333 0.368000000
## [139] 0.370666667 0.373333333 0.376000000 0.378666667 0.381333333 0.384000000
## [145] 0.386666667 0.389333333 0.392000000 0.394666667 0.397333333 0.400000000
## [151] 0.402666667 0.405333333 0.408000000 0.410666667 0.413333333 0.416000000
## [157] 0.418666667 0.421333333 0.424000000 0.426666667 0.429333333 0.432000000
## [163] 0.434666667 0.437333333 0.440000000 0.442666667 0.445333333 0.448000000
## [169] 0.450666667 0.453333333 0.456000000 0.458666667 0.461333333 0.464000000
## [175] 0.466666667 0.469333333 0.472000000 0.474666667 0.477333333 0.480000000
## [181] 0.482666667 0.485333333 0.488000000 0.490666667 0.493333333 0.496000000
## [187] 0.498666667
##
## $spec
## [1] 50949850246 40776827194 26503819948 3479247908 3011402764
## [6] 4831740271 4518314121 3455028331 4093566095 12503235397
## [11] 32725915389 41162613084 1551844848 2286585475 1168946652
## [16] 1190739759 317221075 173078075 1618968655 1448150797
## [21] 1910027538 21417046154 167429670560 13527292168 5942480027
## [26] 1338085555 3246200626 312483063 1119479986 1396165467
## [31] 642126491 2533579929 4772224291 54119857338 57142138559
## [36] 9066691214 174589663 492637233 2028148468 621464605
## [41] 2160489443 1444920232 402867866 2902513883 29186241467
## [46] 81883547676 8774369960 12593248386 3810776094 3412422597
## [51] 416027417 817648369 1655600334 1224767772 2339339403
## [56] 16551483965 84943308419 119197191769 38048539225 6406112094
## [61] 3850810789 1660316541 106750356 525800811 184795889
## [66] 3674740632 3490379125 4893419803 4368743656 4342867488
## [71] 3163326081 2539344685 131707212 319817785 1216323268
## [76] 594980359 448698568 353382502 1283531079 7812860078
## [81] 11242170295 3412348367 5059220848 4977403479 2184829562
## [86] 323812076 770506901 2675860201 8753534188 8578009806
## [91] 46351314015 88229721735 72318375750 78727247165 49459527350
## [96] 6385906837 3325112848 19554550 2081925937 6608964163
## [101] 32328248305 95574966524 98441050966 16543940997 44211243874
## [106] 19799153111 10806393167 4195784702 4079081477 4775670547
## [111] 818573275 5893153075 26124760393 14432167611 8171109650
## [116] 6261715536 8634265442 2313294567 1531634194 400535674
## [121] 935254610 1803956140 513662303 4729494972 16929486245
## [126] 5974534697 7135555487 8216891755 10850432619 4461581232
## [131] 1470226855 925424064 3363878222 1011911300 7527142438
## [136] 37959611658 39906261572 2226320066 10662073971 7583294548
## [141] 35354960488 31819982353 5548390821 4467753586 4978444523
## [146] 2762197001 141495395359 106596963259 42101426500 1690375243
## [151] 29849173419 29214192968 31994589777 7457334219 999400917
## [156] 2058067964 191829913 6219656305 27677105507 22998520288
## [161] 3301684096 8142049596 10180409334 35093696472 23886313416
## [166] 1592900613 4086013528 9659698427 19552249900 118072021371
## [171] 91625890917 22775889275 8347747971 3108316545 4806549253
## [176] 3951973413 10447364921 3124119606 9358347100 9084533804
## [181] 55878043399 100718019378 24901974244 3797193636 2274789065
## [186] 46043178 3359164397
##
## $coh
## NULL
##

```

```

## $phase
## NULL
##
## $kernel
## NULL
##
## $df
## [1] 1.729482
##
## $bandwidth
## [1] 0.0007698004
##
## $n.used
## [1] 375
##
## $orig.n
## [1] 362
##
## $series
## [1] "x"
##
## $snames
## NULL
##
## $method
## [1] "Raw Periodogram"
##
## $taper
## [1] 0.1
##
## $pad
## [1] 0
##
## $detrend
## [1] TRUE
##
## $demean
## [1] FALSE
##
## attr(,"class")
## [1] "spec"

```

Berdasarkan Periodogram Produksi Padi bisa dilihat bahwa ada fluktuasi yang berulang dalam data sehingga bisa disimpulkan adanya pola musiman. Melihat hal ini maka akan dilakukan analisis prediksi yang lebih lanjut dan komplek.

## Model

Clustering Produksi Padi Tujuan klastering adalah mengelompokkan wilayah produksi padi berdasarkan kabupaten dengan pola produksi yang sama.



```
# Filter data produksi terbanyak berdasarkan tahun
grouped_data <- padisumutfilter %>%
  group_by(Tahun) %>%
  summarise(Total_Produksi = sum(Produksi)) %>%
  arrange(desc(Total_Produksi))
grouped_data
```

```
## # A tibble: 13 x 2
##   Tahun      Total_Produksi
##   <date>          <dbl>
## 1 2017-01-01      5136186
## 2 2016-01-01      4609791
## 3 2015-01-01      4026449
## 4 2013-01-01      3727682
## 5 2012-01-01      3715084
## 6 2014-01-01      3631039
## 7 2011-01-01      3607405
## 8 2010-01-01      3582302
## 9 2018-01-01      2108285.
## 10 2019-01-01      2078902.
## 11 2020-01-01      2040500.
## 12 2021-01-01       392937
## 13 2022-01-01      103327.
```

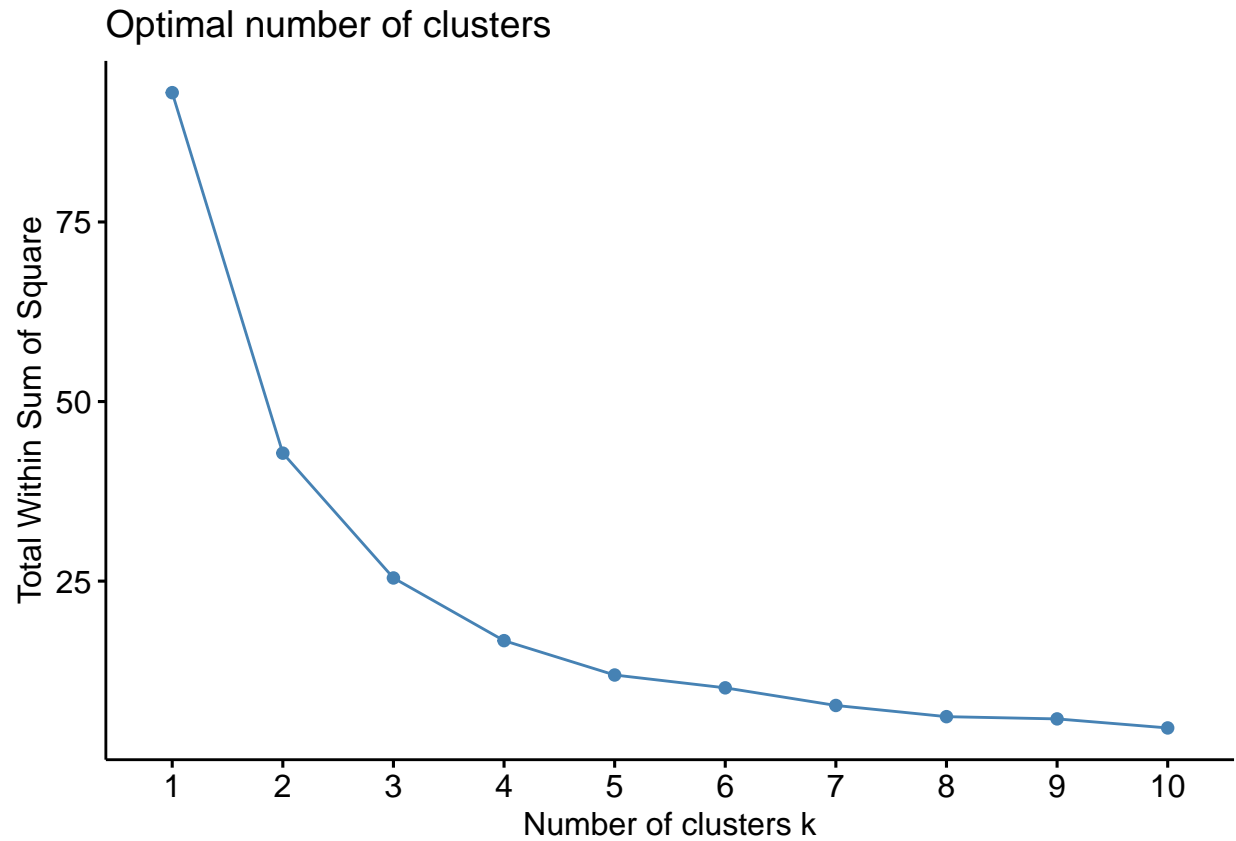
Berdasarkan hasil diatas bisa dilihat jumlah total produksi terbanyak ada pada Tahun 2017, yaitu sebanyak 5136186.0 Ton

```
# Menyiapkan data, untuk pengujian klastering data yang ambil adalah data tahun 2017 saja
datapadi2017 <- padisumutfilter %>%
  filter(Tahun == "2017-01-01")
```

```
# Memilih variabel klastering
clustering_data <- datapadi2017 %>%
  dplyr::select(Produksi, Rata_rata_produksi, Luas_panen)
```

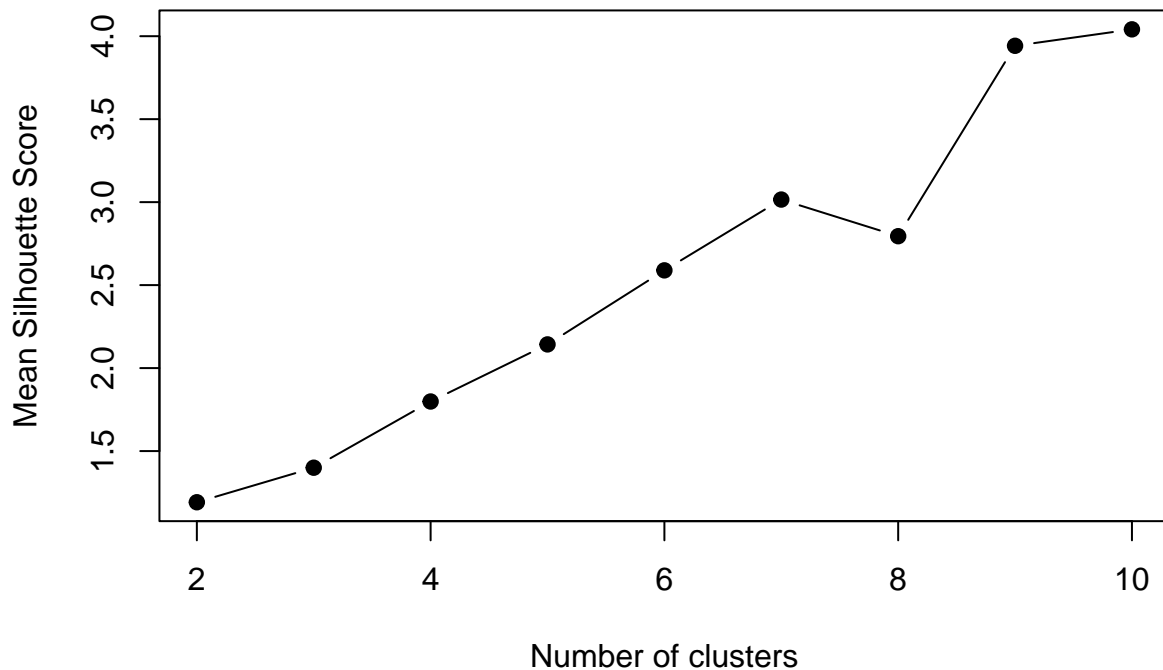
```
# Tentukan skala dari variabel
scale_data <- as.data.frame(scale(clustering_data))
```

```
# Menggunakan Elbow method untuk menentukan jumlah klaster optimal
set.seed(123)
fviz_nbclust(scale_data, kmeans, method = "wss")
```



```
# Menggunakan Silhouette method untuk menentukan klaster optimal
silhouette_scores <- sapply(2:10, function(k) {
  kmeans_result <- kmeans(scale_data, centers = k)
  mean_silhouette <- mean(silhouette(kmeans_result$cluster, dist(scale_data)))
  return(mean_silhouette)
})

# Plot silhouette scores
plot(2:10, silhouette_scores, type = "b", pch = 19, xlab = "Number of clusters", ylab = "Mean Silhouette")
```



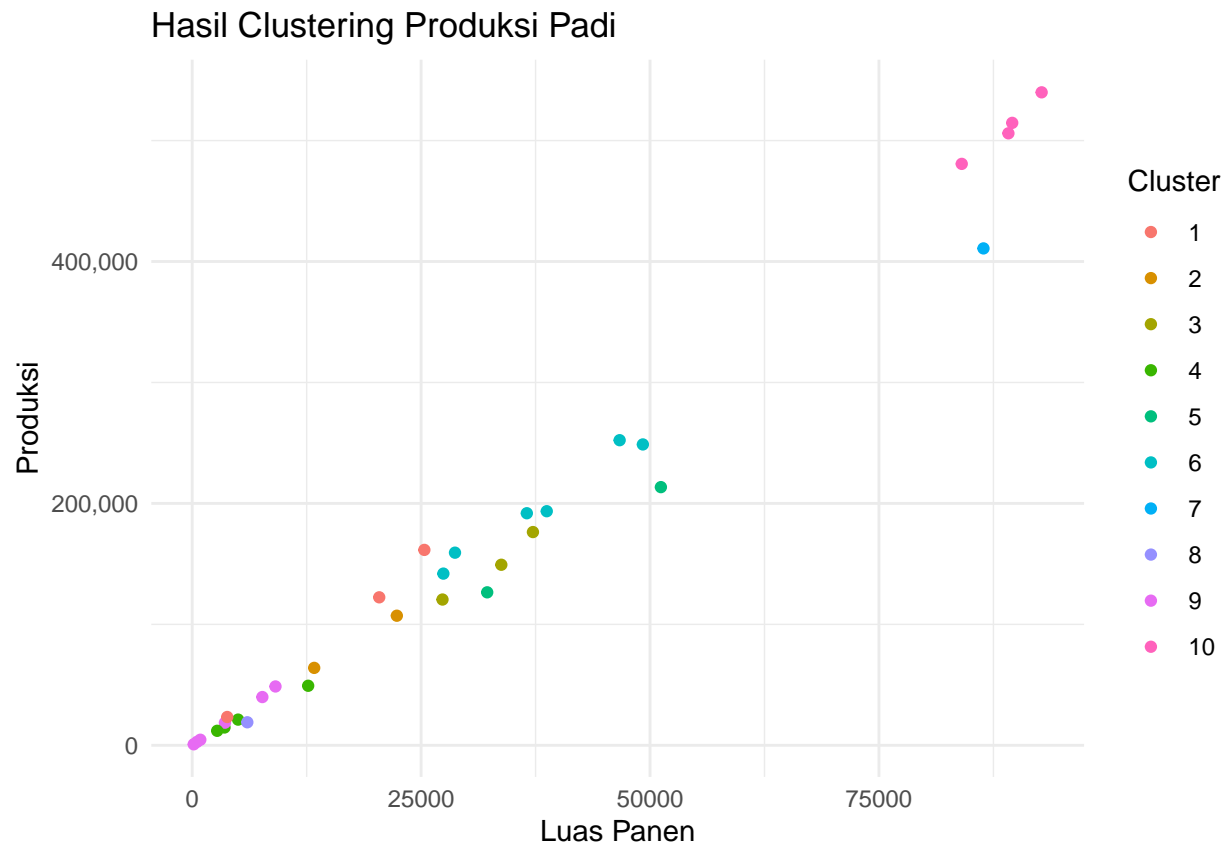
```
# Mencari jumlah klaster terbaik dengan melihat nilai silhouette score
optimal_k <- which.max(silhouette_scores) + 1
cat("Optimal number of clusters:", optimal_k)
```

```
## Optimal number of clusters: 10
```

```
# Klastering dengan K-Means
set.seed(123)
k <- 10 # jumlah klaster
kmeans_result <- kmeans(scale_data, centers = k)
```

```
# tambahkan label ke dalam data frame
datapadi2017$Cluster <- as.factor(kmeans_result$cluster)
```

```
# Plot hasil klaster menggunakan ggplot2
plot_8 <- ggplot(datapadi2017, aes(x = Luas_panen, y = Produksi, color = factor(Cluster))) +
  geom_point() +
  labs(title = "Hasil Clustering Produksi Padi", x = "Luas Panen", y = "Produksi") +
  scale_color_discrete(name = "Cluster") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal()
plot_8
```



```
# Menghitung jumlah klaster per kabupaten/kota
```

```
cluster_counts_per_kabupaten <- datapadi2017 %>%
  group_by(Kabupaten_kota, Cluster) %>%
  count()
```

```
cluster_counts_per_kabupaten
```

```
## # A tibble: 32 x 3
## # Groups:   Kabupaten_kota, Cluster [32]
##   Kabupaten_kota    Cluster     n
##   <chr>           <fct>   <int>
## 1 Asahan          1         1
## 2 Batu Bara       6         1
## 3 Binjai          4         1
## 4 Dairi           6         1
## 5 Deli Serdang    10         1
## 6 Gunungsitoli     9         1
## 7 Humbang Hasundutan 2         1
## 8 Karo            6         1
## 9 Labuanbatu Utara  6         1
## 10 Labuhan Batu    6         1
## # i 22 more rows
```

```
# Hitung jumlah kabupaten setiap klaster
cluster_kabupaten_counts <- datapadi2017 %>%
  filter(Cluster %in% c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)) %>%
  group_by(Cluster) %>%
  summarise(Jumlah_Kabupaten = n())
```

```
cluster_kabupaten_counts
```

```
## # A tibble: 10 x 2
##   Cluster Jumlah_Kabupaten
##   <fct>          <int>
## 1 1          3
## 2 2          2
## 3 3          3
## 4 4          4
## 5 5          2
## 6 6          6
## 7 7          1
## 8 8          1
## 9 9          6
## 10 10         4
```

```
# Menampilkan hasil klaster dengan label nama kabupaten/kota
plot_9 <- ggplot(datapadi2017, aes(x = Luas_panen, y = Produksi, label = Kabupaten_kota)) +
  geom_point(aes(color = factor(Cluster))) +
  geom_text(vjust = -0.5, hjust = 0, size = 3, show.legend = FALSE) +
  labs(title = "Hasil Clustering Produksi Padi", x = "Luas Panen (Ha)", y = "Produksi (Ton)") +
  scale_color_discrete(name = "Cluster") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal()
plot_9
```

Scatter plot showing the relationship between Luas Panen (Ha) on the x-axis and Luas Tanaman (Ha) on the y-axis for various districts in Sumatra. The x-axis ranges from 0 to 75,000 Ha, and the y-axis ranges from 0 to 100,000 Ha. Data points are labeled with district names, showing a positive correlation between the two variables.

District	Luas Panen (Ha)	Luas Tanaman (Ha)
Tanjung Beringin	~500	~500
Merangin	~1000	~1000
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500
Padang Pari Utara	~1500	~1500
Padang Pari Timur	~1500	~1500
Padang Pari Barat	~1500	~1500
Padang Pari Selatan	~1500	~1500

```
## # A tibble: 10 x 2
##   Cluster Avg_Produksi
##   <fct>         <dbl>
## 1 1          102440
## 2 2           85593
## 3 3          148706
## 4 4           24299
## 5 5          169982.
## 6 6          197953.
## 7 7          410837
## 8 8           19042
## 9 9          19256.
## 10 10         510318.
```

```
##      Produksi Rata_rata_produksi  Luas_panen
## 1  -0.35533115          1.5929550 -0.48882742
```

```
## 2 -0.45842562 -0.2823471 -0.44449534
## 3 -0.07220843 -0.6651322 0.06449427
## 4 -0.83351152 -1.1458551 -0.84847743
## 5 0.05798605 -1.3203500 0.36895640
## 6 0.22915742 0.3222924 0.23877208
## 7 1.53189056 -0.3464895 1.89299397
## 8 -0.86568150 -2.5411240 -0.84723336
## 9 -0.86437296 0.3165449 -0.92795025
## 10 2.14065705 1.0149840 1.97682388
```

```
# Definisikan kriteria klaster
low_criteria <- quantile(cluster_avg$Avg_Produksi, 0.25)
high_criteria <- quantile(cluster_avg$Avg_Produksi, 0.75)

# Berikan label untuk masing-masing klaster
cluster_avg_kriteria <- cluster_avg %>%
  mutate(Category = ifelse(Avg_Produksi < low_criteria, "Rendah",
                           ifelse(Avg_Produksi > high_criteria, "Tinggi", "Sedang")))
cluster_avg_kriteria
```

```
## # A tibble: 10 x 3
##   Cluster Avg_Produksi Category
##   <fct>      <dbl> <chr>
## 1 1      102440 Sedang
## 2 2      85593 Sedang
## 3 3     148706 Sedang
## 4 4      24299 Rendah
## 5 5    169982. Sedang
## 6 6    197953. Tinggi
## 7 7     410837 Tinggi
## 8 8      19042 Rendah
## 9 9     19256. Rendah
## 10 10    510318. Tinggi
```

Berdasarkan hasil perhitungan nilai centroid dari klaster, maka bisa disimpulkan bahwa: 1. Klaster 4, 8 dan 9: adalah klaster yang termasuk kategori rendah 2. Klaster 1,2,3 dan 5: adalah klaster yang termasuk kategori sedang 3. Klaster 6, 7 dan 10: adalah klaster yang termasuk kategori tinggi

```
# Menghitung jumlah klaster per kelompok kabupaten/kota
kabupaten_per_cluster <- datapadi2017 %>%
  group_by(Cluster) %>%
  summarise(Kabupaten_kota = list(unique(Kabupaten_kota)))

# Menampilkan nama kabupaten/kota dari setiap klaster
for (i in 1:10) {
  cat(paste("Klaster", i, ":", kabupaten_per_cluster$Kabupaten_kota[[i]], "\n"))
}
```

```
## Klaster 1 : Asahan
## Klaster 1 : Pematangsiantar
## Klaster 1 : Toba Samosir
## Klaster 2 : Humbang Hasundutan
## Klaster 2 : Nias
```

```

## Klaster 3 : Nias Selatan
## Klaster 3 : Tapanuli Tengah
## Klaster 3 : Tapanuli Utara
## Klaster 4 : Binjai
## Klaster 4 : Medan
## Klaster 4 : Nias Barat
## Klaster 4 : Nias Utara
## Klaster 5 : Padang Lawas
## Klaster 5 : Padang Lawas Utara
## Klaster 6 : Batu Bara
## Klaster 6 : Dairi
## Klaster 6 : Karo
## Klaster 6 : Labuanbatu Utara
## Klaster 6 : Labuhan Batu
## Klaster 6 : Tapanuli Selatan
## Klaster 7 : Mandailing Natal
## Klaster 8 : Pakpak Bharat
## Klaster 9 : Gunungsitoli
## Klaster 9 : Labuhanbatu Selatan
## Klaster 9 : Padangsidempuan
## Klaster 9 : Samosir
## Klaster 9 : Tanjungbalai
## Klaster 9 : Tebing Tinggi
## Klaster 10 : Deli Serdang
## Klaster 10 : Langkat
## Klaster 10 : Serdang Bedagai
## Klaster 10 : Simalungun

```

## Evaluasi Model

Evaluasi Clustering dengan Inertia dan mengukur validitas dari model prediksi

```

# Jumlah klaster yang diuji
k_values <- 2:10

# Hitung nilai Inertia untuk setiap klaster
inertia_values <- sapply(k_values, function(k) {
  kmeans_result <- kmeans(clustering_data, centers = k)
  kmeans_result$tot.withinss
})

inertia_df <- data.frame(K = k_values, Inertia = inertia_values)
inertia_df

```

```

##      K      Inertia
## 1  2 189904583648
## 2  3  43838059610
## 3  4  22711889738
## 4  5  18463602951
## 5  6  16146884708
## 6  7  13601464875
## 7  8  11602908557
## 8  9  10946905057

```



```
## 9 10 4138350808
```

Berdasarkan hasil evaluasi dengan Inertia, didapatkan bahwa jumlah klaster yang tepat untuk pengujian klaster yang dilakukan adalah 10, karena nilai Inertia yang paling rendah

```
akurasi_klastering <- (kmeans_result$betweenss/kmeans_result$totss)*100  
akurasi_klastering
```

```
## [1] 96.02365
```

Akurasi klaster bisa diketahui dengan menghitung rasio dari jumlah kuadrat antar klaster dengan jumlah kuadrat total. Sehingga nilai validitas dari model klastering adalah 96,02