



**TECNOLÓGICO
NACIONAL DE MÉXICO**



SE

SECRETARÍA DE EDUCACIÓN

GOBIERNO DE BAJA CALIFORNIA



**Tecnológico Nacional de México
Instituto Tecnológico de Tijuana**

Subdirección Académica

Departamento de Sistemas y Computación

SEMESTRE:

Febrero-Julio 2021

CARRERA:

Ingeniería en Sistemas Computacionales e Ingeniería en Tecnologías de la
Información y Comunicaciones

MATERIA Y SERIE:

Minería de Datos

BDD-1703 TI9A

UNIDAD A EVALUAR:

Unidad IV

NOMBRE DEL TRABAJO:

Evaluación U4

Evaluación de la Unidad 4

NOMBRE Y NÚMERO DE CONTROL DE LOS INTEGRANTES:

Rodriguez Medrano Marco Antonio 17210635

NOMBRE DEL DOCENTE:

José Christian Romero Hernández

17 de Junio de 2021



Instrucciones

Desarrolle el siguiente problema con R y RStudio para la extracción de conocimiento que el problema requiere. Implementar el modelo de agrupación K-Means con el conjunto de datos Iris.csv que se encuentra en <https://github.com/jcromerohdz/iris> utilizando el método kmeans() en R. Una vez que se obtenga el modelo de agrupamiento hacer el análisis.

de visualización de datos correspondiente. Al finalizar el desarrollo explicar detalladamente en qué consiste el modelo de agrupación K-Means y cuáles fueron sus observaciones en el análisis de visualización de datos.

Instrucciones de evaluación

- Tiempo de entrega 3 días
- Al terminar poner el código y la explicación en el branch correspondiente de su github así mismo realizar su explicación de la solución en su google drive.
- Finalmente defender su desarrollo en un video de 8-10 min el cual servirá para dar su calificación, este video debe subirse a youtube para ser compartido por un link público.

```
# Implementación de nuestro espacio de trabajo
```

```
getwd()
```

```
setwd("C:/Users/DELL/Desktop/Semestre9/MineriaDatos/EvaluacionU4")
```

```
getwd()
```

```
#Importación del Dataset iris
```

```
dataset = read.csv('iris.csv')
```

```
#Seleccionando los datos que usamos
```

```
dataset = dataset[1:4]
```

```
#Instalacion de la librería cluster que usaremos para encontrar  
los clusters
```

```
#install.packages("cluster")
```

```
library(cluster)
```

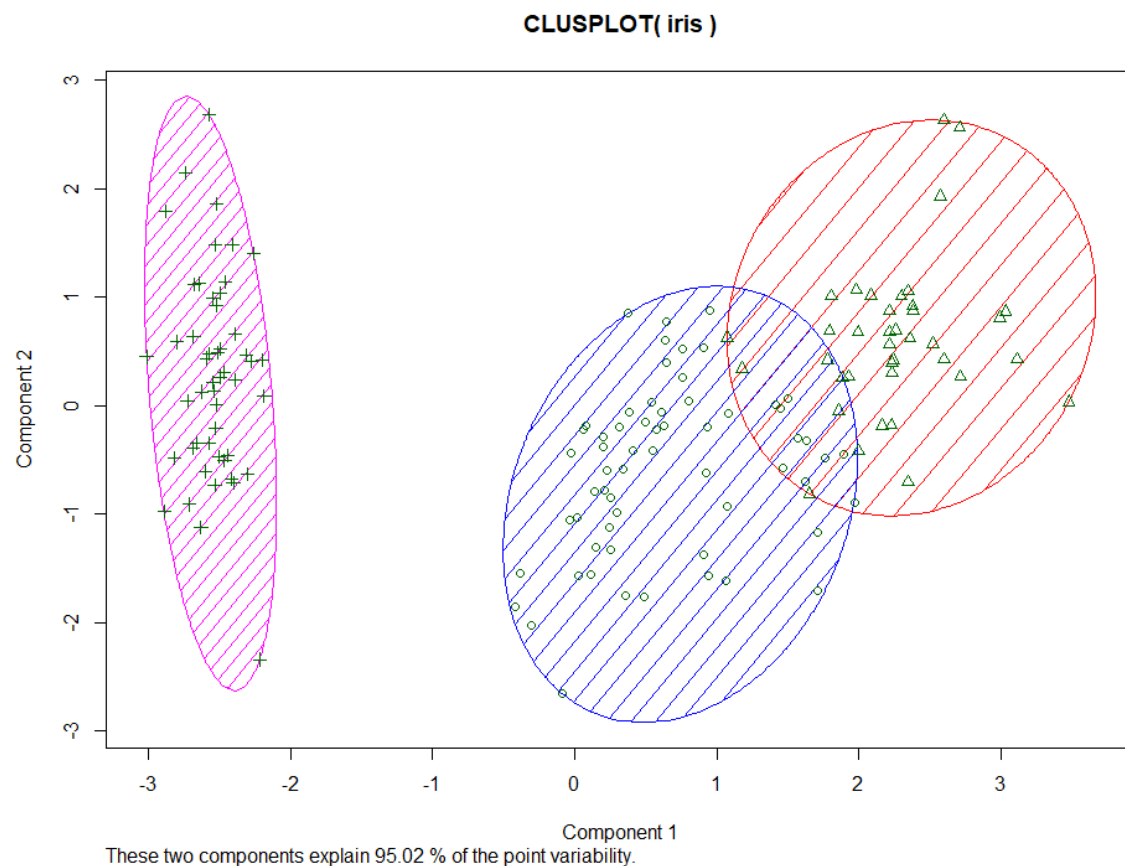
```
# Insertamos nuestra semilla de aleatoriedad, en la función  
kmeans, esto es necesario para establecer el centro,
```

```
# cuál es la cantidad de grupos que queremos y se agruparán
```

```
set.seed(101)
```

```
irisCluster <- kmeans(dataset[,1:4], center=3, nstart=20)
```

```
irisCluster
```





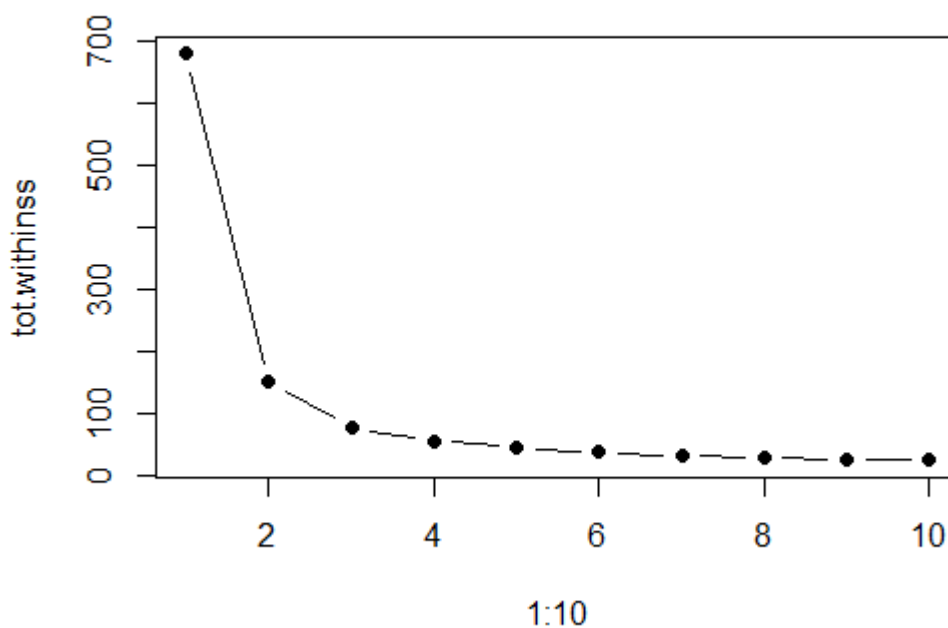
#Aplicaremos el método del codo para ver mejor la implementación de los datos.

```
tot.withinss <- vector(mode="character", length=10)
for (i in 1:10){
  irisCluster <- kmeans(dataset[,1:4], center=i, nstart=20)
  tot.withinss[i] <- irisCluster$tot.withinss
}
```

#Visualizamos el método.

```
plot(1:10, tot.withinss, type="b", pch=19)
```

#Resultado



Quiero darle crédito a mi empresa; miembro del equipo Ramon Loaiza Chavez,

#porque me ayudó a resolver un problema en el programa, sin su ayuda no podría entregar esta evaluación

Conclusión

El método de k means es para mi el mejor método de agrupación de datos que puedas implementar en tus programas en R una por su simpleza, dos por su fácil implementación y la tercera que considero que es la más importante es que nos permite hacer una análisis muy completo a los datos y así tomar decisiones.