**Tecnológico Nacional De México**

**Instituto Tecnológico De Tijuana**

**Subdirección Académica**

**Departamento de Sistemas y Computación**

**Semestre Enero - Junio 2022**

**Ingeniería Informática**

**Mineria De Datos**

**Práctica Evaluatoria**

**Unidad 4**

**Díaz Ruiz Uriel No.18210839**

**Perez Ortega Victoria Valeria No.18210718**

**JOSE CHRISTIAN ROMERO HERNANDEZ**

**Tijuana, B.C. a 01 de Junio de 2022.**

**TECNOLÓGICO NACIONAL DE MÉXICO**

***INSTITUTO TECNOLÓGICO DE TIJUANA***

**SUBDIRECCIÓN ACADÉMICA**

Departamento de Sistemas y Computación

**EXAMEN**

Carrera: Ingeniería En Sistemas Computacionales/ Tecnologías de la información/ Informática Período: **Febrero-Junio 2022** Materia: Minería de datos Grupo: Salón:1 Unidad (es) a evaluar: Unidad 4 Tipo de examen:Práctico Fecha: Catedrático: José Christian Romero Hernandez Firma del maestro: Calificación:

Alumno:Diaz Ruiz Uriel y Perez Ortega Victoria Valeria No. Control:18210839, 18210718 **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 6-8 min el cual servirá para dar su calificación, este video debe subirse a youtube para ser compartido por un link público.

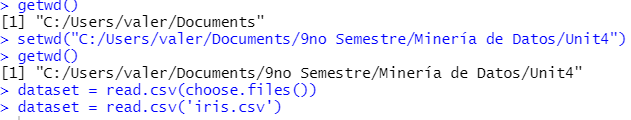
**Instalación de paquetes:**

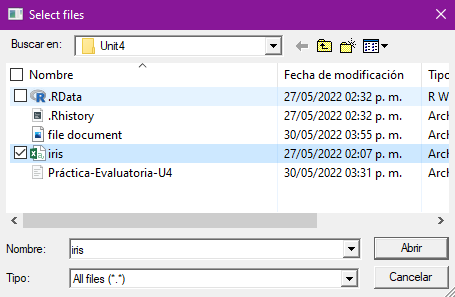
| install.packages("cluster")  library(cluster) |
| --- |

**Se carga el archivo de datos iris.csv:**

getwd()

| setwd("C:/Users/valer/Documents/9no Semestre/Minería de Datos/Unit4")  getwd() dataset = read.csv(choose.files()) dataset = read.csv('iris.csv') |
| --- |





**Aquí hacemos referencia a que los datos se van a relacionar con las primeras 4 columnas:**

| dataset = dataset[1:4] |
| --- |



**Aquí decimos cuántos son los clusters que se van a utilizar, utilizando el método elbow, se crea un vector que nos ayudará a guardar la suma de cuadrados de cada grupo de clusters:**

| set.seed(6)  wcss = vector() |
| --- |

**Con ayuda del ciclo for se le va a decir cuántas veces queremos que se guarde la información y al final obtendremos el número de clusters que son 3, lo cual nos mostrará con una una curva en la gráfica:**

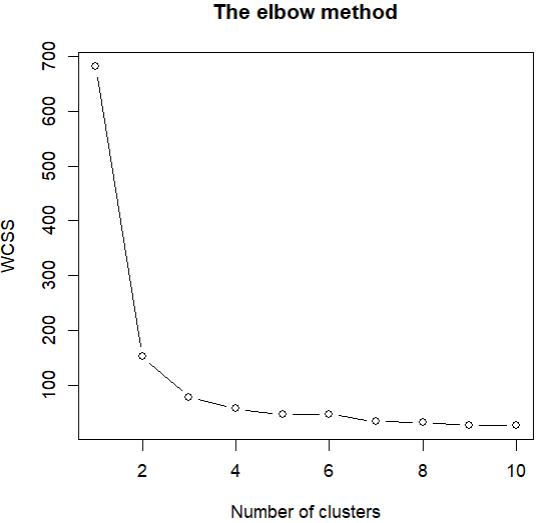
| for (i in 1:10) wcss[i] = sum(kmeans(dataset, i)$withinss) plot(1:10,wcss,type = 'b', main = paste('The elbow method'), xlab = 'Number of clusters', ylab = 'WCSS') |
| --- |

**Ahora debemos de ejecutar nuestro modelo k-mean en que está instalado en el paquete base de R-Studio, Ahí debemos de agregar nuestra semilla de nuestra aleatoriedad, en la función de nuestro kmeans, el cuál es de gran necesidad tener establecido el centro el cual es el número de todos los grupos que queremos y se deben de agrupar:**

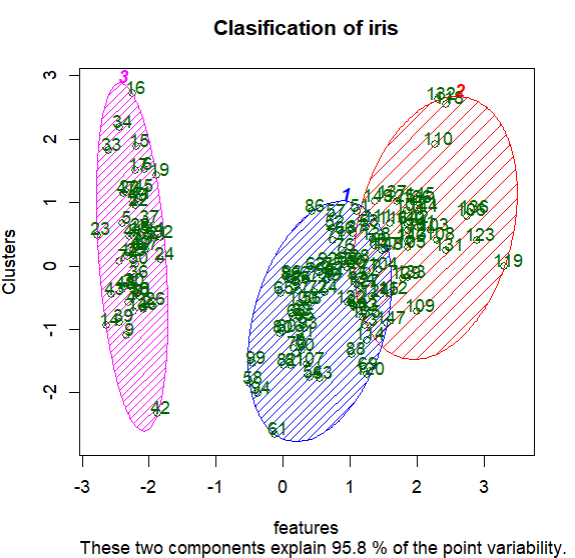
**Debemos de aplicar el método del codo para que se pueda observar la implementación de los datos y dónde está el punto de quiebre de la gráfica obtenida:**

| set.seed(29)  kmeans = kmeans(x = dataset, centers = 3)  y\_kmeans = kmeans$cluster |
| --- |

| library(cluster)  clusplot(dataset,  y\_kmeans,  lines = 0,  shade = TRUE,  color = TRUE,  labels = 2,  plotchar = FALSE,  span = TRUE,  main = paste('Clasification of iris'),  xlab = 'features',  ylab = 'Clusters') |
| --- |



| set.seed(29) kmeans = kmeans(x = dataset, centers = 3)  y\_kmeans = kmeans$cluster library(cluster) clusplot(dataset, y\_kmeans, lines = 0,  shade = TRUE,  color = TRUE,  labels = 2,  plotchar = FALSE,  span = TRUE, main = paste('Clasification of iris'),  xlab = 'features', ylab = 'Clusters') |
| --- |



**URL:** [**https://youtu.be/OnRG\_e6nfyA**](https://youtu.be/OnRG_e6nfyA)