

---

## PRACTICA 2

### Objetivo General

Conocer mas a detalle el proceso de KDD sobre una base de datos del aprendizaje no supervisado. Así como la implementación del primer clasificador del aprendizaje no supervisado.

### Objetivos Particulares:

- Implementar el clasificador del aprendizaje no supervisado llamado K-Medias en C.
- Incorporar las fases de *integración y recopilación, selección, limpieza y transformación* sobre una base de datos artificial (empresas de clase media) y una base datos del aprendizaje supervisado (temblores) a fin de obtener una vista minable de los datos.

### Documentos para entregar

Esta práctica se debe realizar de forma individual, los documentos a entregar son los siguientes:

- Implementación del algoritmo K-Medias y un script en donde se realice el procesamiento de los datos crudos en C (70 % de la calificación).
- Informe individual de cómo se abordaron cada uno de los problemas planteados (30 % de la calificación). El informe debe ser a mano y contener:
  - Caratula, con nombre de la materia, nombre propio (empezando por apellidos), matricula, profesor y fecha de entrega.
  - Introducción, que es la Clase Graphics y para que se ocupa.
  - Una sección de Desarrollo en el cual me expliquen usando parcialidades de código y algún otro recurso (pantallazos, esquemas, etc) como se abordaron cada una de las partes que conforman la práctica.
  - Conclusiones generadas de la Practica

### Plazo de entrega

La hora y fecha límite para enviar la practica será el miércoles 02 de noviembre del 2022 antes de las 10:00 AM. Al correo opelo1209@gmail.com

**Nota. No se recibirá ninguna práctica fuera de ese horario, sin ninguna excepción.**

**Especificaciones de las partes que conforman la práctica:****Parte I:**

Realizar un programa (Transformación\_ASCII.c) para generar una base de datos de empresas de clase media (BD\_ECM.dat) en un formato de tabla con sus respectivos valores de código ASCII. Los atributos de la BD\_ECM se enlistan a continuación (recordad que no hay acentos, pero si la letra ñ):

- Costumbres Sociales:
  - Simbolico
  - Ritual
  - Ceremonial
  - Relato
  - Modelos
- Estatus:
  - Trascendencia
  - Prestigio
  - Rango
- Cooperación:
  - Imitación
  - Seguridad
- Psicológicos:
  - Egoismo
  - Confianza
  - Indiferencia
  - Precaución
  - Premonicion
  - Fe
- Viscerales:
  - Dolor
  - Hambre
  - Sueño
  - Agotamiento
-

- Emociones:
  - Positivas
  - Negativas
- Conocimientos:
  - Habituales
  - Destrezas
  - Limitados
  - Trucos
  - Engaños

En esta parte no se necesitan generar los 28,152,495, los cuales que se obtienen al realizar la combinación de cada uno de los valores de cada atributo y después la combinación con cada atributo, esto es dejar que los entrevistados escojan más de un valor de los atributos. En este caso solo se pide que obtengan las 7200 combinaciones tomando un solo valor de cada atributo y combinarlo con los demás atributos. La salida de esta parte del código se puede observar en la Figura 1. Generar las conclusiones adecuadas para esta parte de la práctica.

```
G:\Documentos_Benja\Trabajo_UAM\Trimestre_20-01\Materia - Aprendizaje Maquina\Practicas\Practica3\Transformacion_ASCII.exe
#####
TRANSFORMACION DE LA BASE DE DATOS DE EMPRESAS DE CLASE MEDIA

La transformacion de los atributo de Costumbres Sociales son:
  El codigo ASCII de Simbolico es: 929
  El codigo ASCII de Ritual es: 625
  El codigo ASCII de Ceremonial es: 1023
  El codigo ASCII de Relatos es: 730
  El codigo ASCII de Modelos es: 723

La transformacion de los atributo de Estatus son:
  El codigo ASCII de Trascendencia es: 1332
  El codigo ASCII de Prestigio es: 950
  El codigo ASCII de Rango es: 503

La transformacion de los atributo de Cooperacion son:
  El codigo ASCII de Imitacion es: 925
  El codigo ASCII de Seguridad es: 920

La transformacion de los atributo de Psicologicos son:
  El codigo ASCII de Egoismo es: 723
  El codigo ASCII de Confianza es: 921
  El codigo ASCII de Indiferencia es: 1217
  El codigo ASCII de Precaucion es: 909
  El codigo ASCII de Premonicion es: 1155
  El codigo ASCII de Fe es: 171

La transformacion de los atributo de Viscerales son:
  El codigo ASCII de Dolor es: 512
  El codigo ASCII de Hambre es: 591
  El codigo ASCII de Susto es: 397
  El codigo ASCII de Agotamiento es: 1144

La transformacion de los atributo de Emociones son:
  El codigo ASCII de Positivas es: 962
  El codigo ASCII de Negativas es: 930

La transformacion de los atributo de Conocimientos son:
  El codigo ASCII de Habituales es: 1026
  El codigo ASCII de Destrezas es: 949
  El codigo ASCII de Limitados es: 934
  El codigo ASCII de Trucos es: 640
  El codigo ASCII de Engaños es: 590

Archivo BD_ASCII.dat generado con exito
Se generaron 7200 registros
GENERACION DE LA BASE DE DATOS FINALIZADA
#####
-----
Process exited after 0.5962 seconds with return value 0
```

**Figura 1.** Salida de la primera parte de la Practica.

**Parte II:**

Desarrollar la implementación del clasificador del aprendizaje no supervisado llamado K-Medias.c y corroborar su funcionamiento con el ejemplo de la presentación de la Semana 2 mediante un script llamado Prueba\_ParteII.c. Una posible salida del programa que desarrollen sería la siguiente:

Parte II. Ejemplo de Clase

Los grupos generados por el algoritmo K-Medias fueron los siguientes:

6	-5	2
2	3	1
6	-2	2
-2	4	1

Probar con varias inicializaciones de los clúster (generación aleatoria del programa, tomar de base la clase de la Semana 3), ¿Cambia el resultado final del programa, explicar por qué si o no?

**Parte III:**

En base a la implementación anterior, modifique el algoritmo para que pueda manejar el criterio de paro en base a un número de iteraciones.

Una vez que se tenga el clasificador no supervisado K-Medias.c con el criterio de paro igual al número de iteraciones, probar esta implementación con la BD\_ECM utilizando los siguientes valores de iteración.

Iteraciones = {2,3,4,5} y  $K = 2$

Reportar cuantos y cuales ejemplos (realizar un pequeño programa Prueba\_ParteIII para reportar esta parte) se agruparon en cada grupo (2 grupos ya que  $K$  la fijamos en 2) para cada uno de los valores de las iteraciones.