



ML LP1: CARGA DE DATOS

Desarrolle un sistema que permita cargar archivos de datos de texto plano, cuyos valores este separador por un delimitador (por ejemplo coma), los cuales podrian contener algo como:

Archivo iris.csv

```
5.1,3.5,1.4,0.2,Iris-setosa  
4.9,3.0,1.4,0.2,Iris-setosa  
6.6,2.9,4.6,1.3,Iris-versicolor  
5.2,2.7,3.9,1.4,Iris-versicolor  
5.8,2.8,5.1,2.4,Iris-virginica  
6.4,3.2,5.3,2.3,Iris-virginica
```

Archivos zoo.csv

```
aardvark,1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1  
antelope,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1  
bass,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4  
bear,1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1  
boar,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
```

El sistema debe tener la siguiente funcionalidad:

1. Permitir especificar el archivo de texto plano a cargar
2. Permitir especificar el delimitador o separador de atributos
3. Automáticamente debe encontrar el numero de atributos (Columnas separador por el delimitador de atributos)
4. Automáticamente encontrar el numero de patrones (renglones con datos)
5. Cargar los atributos en una matriz.
6. Automáticamente definir el tipo de dato y medida.
7. Seleccionar un subconjunto de atributos para generar un vector (ya sea para generar el vector de entrada X y el vector de salida Y)
8. Seleccionar un subconjunto de renglones para generar otra matriz
9. Guardar en un archivo txt plano las matrices reducidas generadas por los dos puntos anteriores.



Para corroborar algunos de los resultados obtenidos puede revisar los archivos de información, los cuales tienen información como:

Informacion del archivo iris

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica

Informacion de la base de datos de Zoo

Attribute Information: (name of attribute and type of value domain)

1. animal name: Unique for each instance
2. hair Boolean
3. feathers Boolean
4. eggs Boolean
5. milk Boolean
6. airborne Boolean
7. aquatic Boolean
8. predator Boolean
9. toothed Boolean
10. backbone Boolean
11. breathes Boolean
12. venomous Boolean
13. fins Boolean
14. legs Numeric (set of values: {0,2,4,5,6,8})
15. tail Boolean
16. domestic Boolean
17. catsize Boolean
18. type Numeric (integer values in range [1,7])