

Jayder Ochoa

Antonio Carmona

Juan Esteban Cardona

Entrega #2:

Analizando el problema que nos planteaban, el cómo podíamos predecir si a una persona le puede ir bien en las pruebas saber pro, y el con que algoritmos podíamos trabajar para poder encontrar la solución a este problema. Nosotros nos encontramos con que hay 2 algoritmos principales con los cuales podemos trabajar y clasificar bien los datasets y los datos de las tablas, al fin concluimos que vamos a usar 2 algoritmos principales: el algoritmo CART, y el algoritmo de C4.5 con lo cual se generan los árboles de decisiones.



Algorithms details

	Splitting criterion	Pruning criterion	Other features
CART	<ul style="list-style-type: none">GiniTwoing	Cross-validation post-pruning	<ul style="list-style-type: none">Regression/Classif.Nominal/numeric attributesMissing valuesOblique splitsNominal splits grouping
ID3	Information Gain (IG)	Pre-pruning.	<ul style="list-style-type: none">ClassificationNominal attributes
C4.5	<ul style="list-style-type: none">Information Gain (IG)Information Gain Ratio (IGR)	Statistical based post-pruning	<ul style="list-style-type: none">ClassificationNominal/numeric attributesMissing valuesRule generatorMultiple nodes split

Pasos a tener en cuenta para desarrollar el funcionamiento del árbol de decisiones:

- Organizar las preguntas según el Gini y su importancia.
- Como calcular el Gini.
- Identificar los tipos de datos.
- Calcular la exactitud del árbol.
- Método para cada persona con sus resultados.

Con esto arrancaríamos con la parte que sigue del proyecto, mas adelante se vera la posibilidad de agregar más métodos o crear unos distintos para poder mejorar el funcionamiento del programa y así elaborar un mejor trabajo.

```
Codigoentrega2.py > ...
1  import pandas as pd
2  import numpy as np
3
4  file = pd.read_csv('lite.csv', sep=';')
5
6
7  def separador(x):
8      a = np.matriz(x.values)
9      for i in range(len(a)):
10         s = a[i, 18]
11         s = s[-4:]
12         p = a[i, 19]
13         p = p//10
14         a[i, 18] = int(p)-int(s)
15         if a[i, 1] == 'no':
16             a[i, 1] == 00
17     a1 = np.delete(a, [0, 11, 12, 22, 24, 48, 50, 57, 58, 62, 64], 1)
18     return a1
19
20
21 print(separador(file)[: , -1].transpose().tolist[0])
22 print(separador(file)[: , :-1].tolist())
23
24 x = separador(file)[: , :-1].tolist()
25 y = separador(file)[: , -1].transpose().tolist[0]
26
```