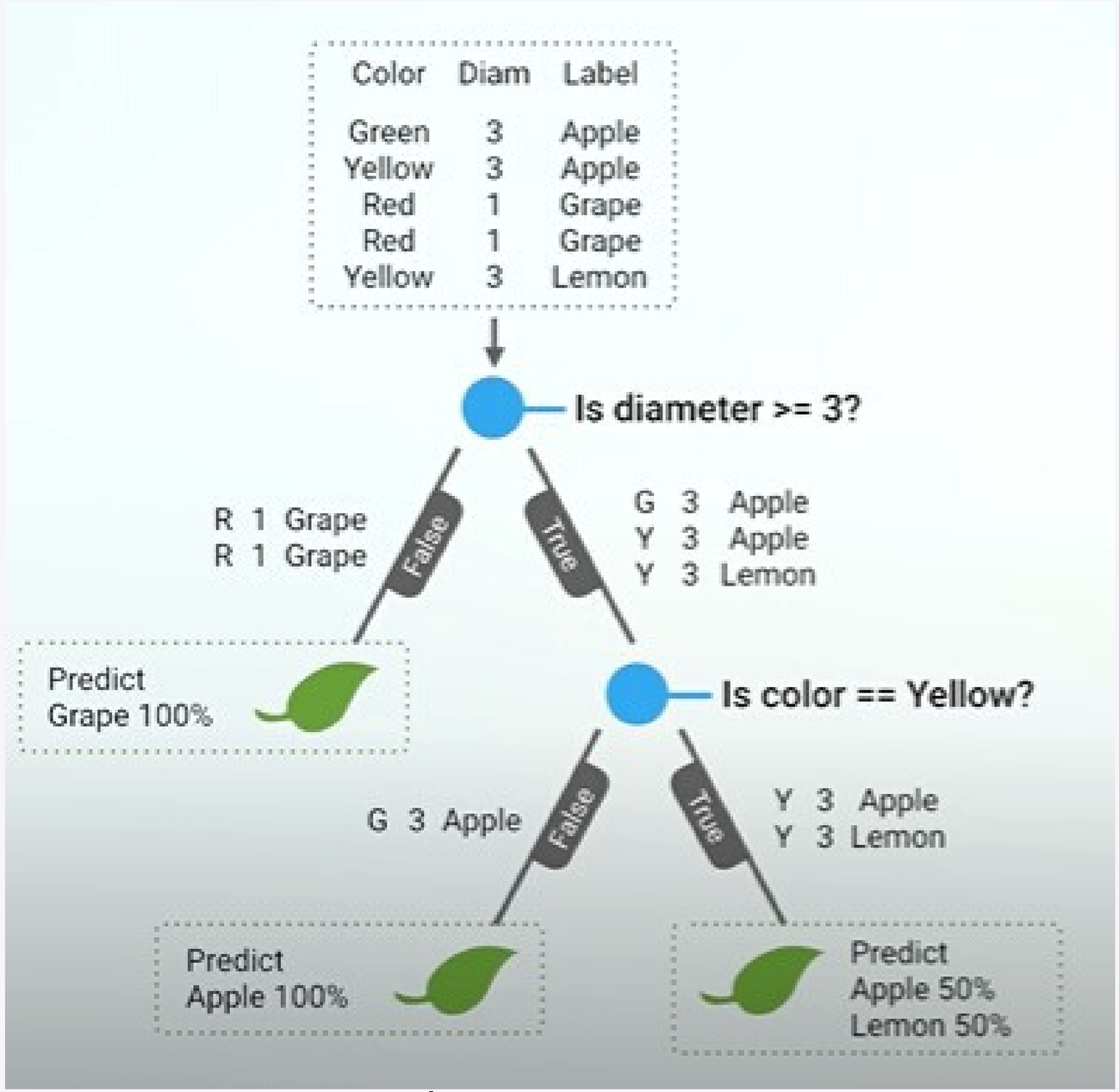
ANÁLISIS PREDICTIVO PARA LAS PRUEBAS SABER PRO

Luis Fernando Vargas Agudelo
Tomás Bedoya Henao

Medellín, mayo 19 / 2020

¿Cómo funciona un árbol de decisión CART?



Gráfica 1. Árbol de decisión

Árbol de decisión CART

• Se agrega un nodo raíz al árbol y este recibe todo el set de entrenamiento.

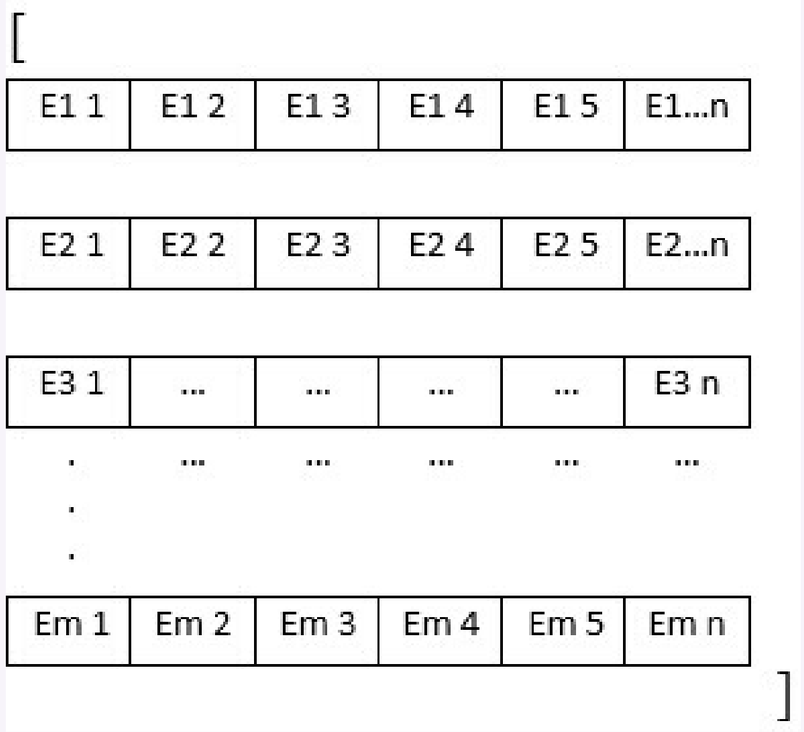
• Se realiza una pregunta sobre una característica en particular con el objetivo de partir el dataset en 2: las filas que retornan verdadero y las que retornan falso; generando así los siguientes nodos.

•¿Cómo son las preguntas?
- == si es texto
- >= si es número

•Se escogen las preguntas indicadas según el valor de la impureza de Gini y la ganancia de información.

•El árbol se construye recursivamente.

Estructura de datos diseñada



Gráfica 2. Estructura lista de listas

Lista de listas

La lista de listas tiene una lista principal como vector unidimensional donde cada elemento de esa lista corresponde a una lista secundaria con todos los atributos del estudiante.

OPERACIONES

Carga_datos

Abre los archivos de entrenamiento, construye una lista de listas a partir de un CSV y retira las columnas innecesarias.

- Dataset de entrenamiento
- Dataset de prueba
- Cabeceras

Conteo

Cuenta la cantidad de valores iguales que existen en una columna.

- Cantidad de valores

Division

Subdivide los datos en aquellos que sí cumplen con la pregunta y aquellos que no.

- Dataset verdadero
- Dataset falso

OPERACIONES

Gini

Calcula la impureza en una lista de filas.
Return: impureza.

Info_gain

Calcula la incertidumbre del dataset y qué tanta información se obtiene al realizar particiones.
Return: info_gain

Preguntas

Genera las preguntas que ayudan a partir el dataset.
Estas preguntas se formulan con ayuda del info_gain.
Return: preguntas

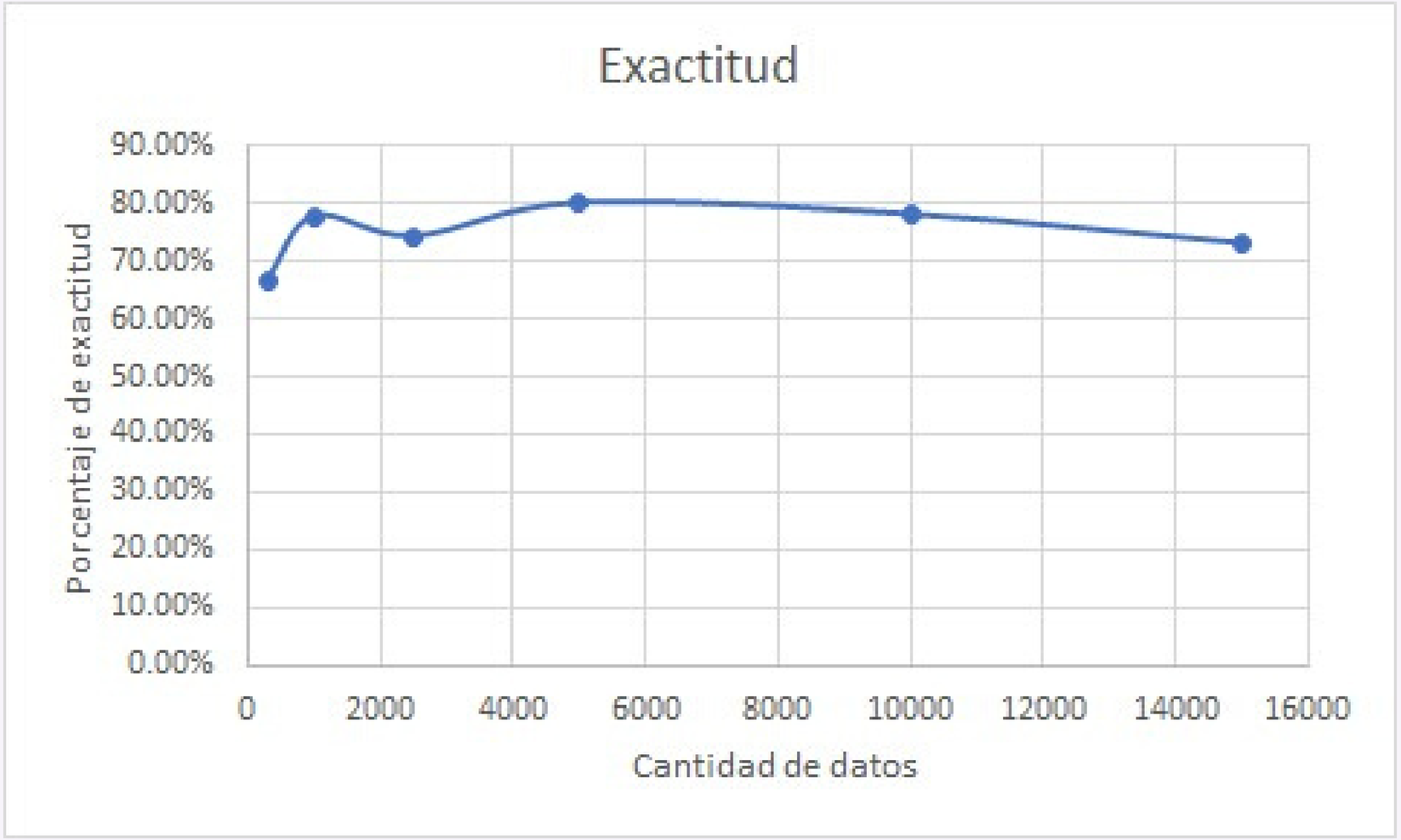
OTRAS OPERACIONES

- Arbol
- Clasificar
- Dibujar arbol
- Exactitud
- Nodo decision
- Hoja
- Mejor division
- Tipo dato
- Valores unicos

Gráfica tiempo

Gráfica 3. Tiempo de ejecución

Gráfica exactitud

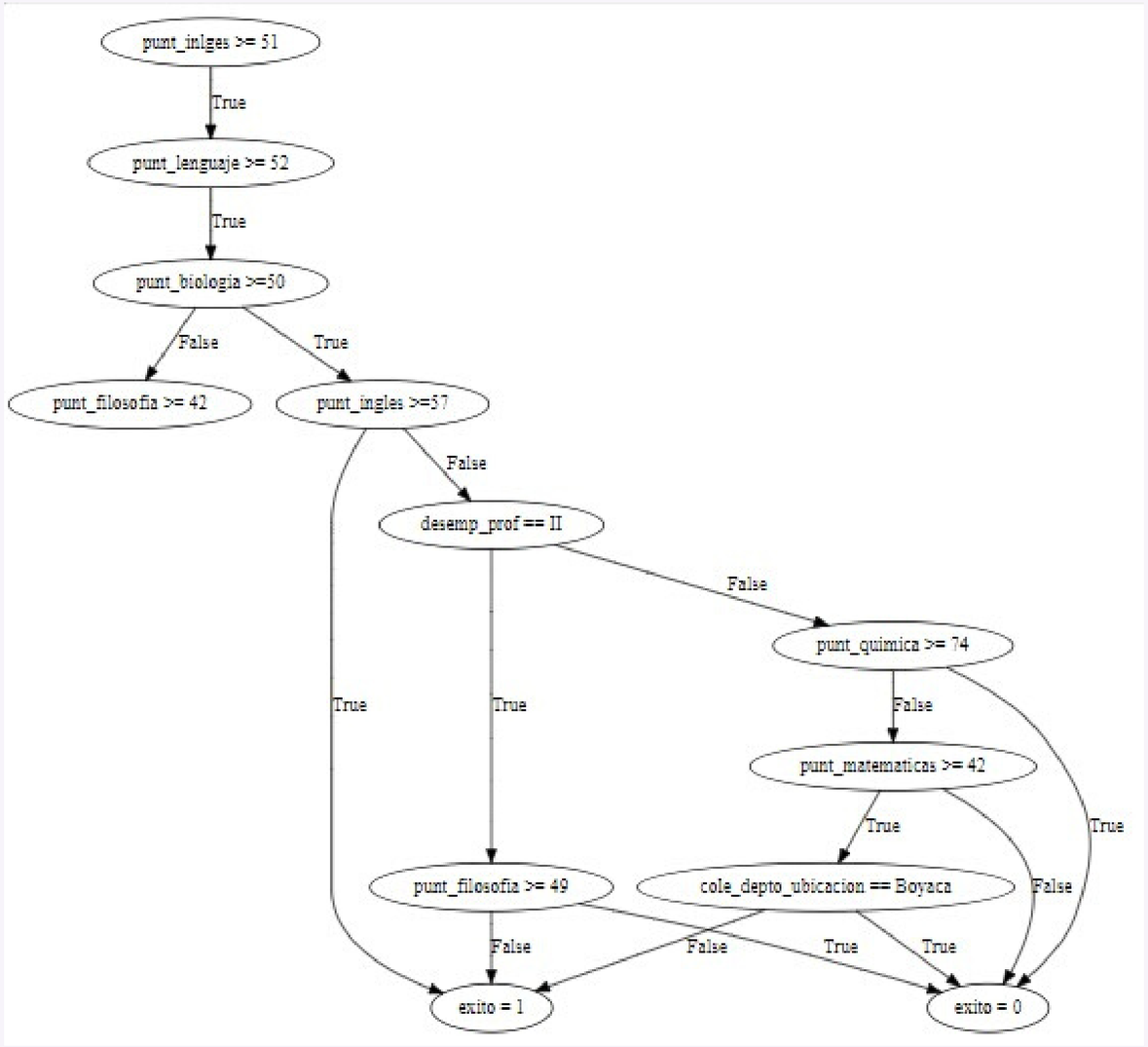


Software desarrollado

```
Exactitud: 81.0%
Tarda 13.14 segundos en ejecutar
Is punt_ingles >= 51?
--> True:
  Is punt_lenguaje >= 52?
  --> True:
    Is punt_biologia >= 50?
    --> True:
      Is punt_ingles >= 57?
      --> True:
        Predict {'1': 103}
      --> False:
        Is desemp_prof == II?
        --> True:
          Is punt_filosofia >= 49?
          --> True:
            Predict {'0': 3}
          --> False:
            Predict {'1': 1}
        --> False:
          Is punt_quimica >= 74?
          --> True:
            Predict {'0': 1}
          --> False:
            Is punt_matematicas >= 42?
            --> True:
              Is cole_depto_ubicacion == BOYACA?
              --> True:
                Predict {'0': 1}
              --> False:
```

Gráfica 5.
Software
desarrollado

Árbol



Gráfica 6. Nodos iniciales del árbol

¡Muchas gracias!