



Factores que Inciden Sobre Completar o No el Grado de Bachillerato en la UPR-Cayey



José Noel Caraballo
Irmannette Torres-Lugo
Universidad de Puerto Rico en Cayey

Temas a Discutirse

- Importancia de la investigación institucional y las decisiones sustentadas por datos.
- Criterios de admisión a la UPR como objeto de estudio.
- Conceptos básicos de la división recursiva.
- Construcción de Árboles
 - Predictores
 - Podar
 - Error de Clasificación
- Prediciendo el éxito de completar un grado
 - Hallazgos
- Discusión y Preguntas

La Investigación Institucional y las Decisiones Sustentadas por Datos

- Cultura de Evaluación y Avalúo Institucional (Meta 2, Díez Década):
 - “Generar y recopilar de forma sistemática y sistematizada la información relevante al comportamiento institucional que permita a todas sus unidades una planificación eficiente y un procesos eficiente de toma de decisiones” (Objetivo 2).
 - “Incentivar la investigación sobre los comportamientos institucionales tanto por los sectores administrativos-gerencias, como por los académicos, en tanto la Universidad es también objeto de estudio” (Objetivo 3).
 - “Auspiciar la discusión informada de los datos institucionales en los foros pertinentes del organigrama universitario y su utilización en la toma ponderada de decisiones institucionales” (Objetivo 4).
- Las políticas respecto a las admisiones de estudiantes procedentes de escuela superior son de interés para la UPRC, particularmente cuando se han desarrollado criterios sistémicos uniformes de admisión a diversos recintos, cada uno con sus particulares y grupos de servicio.

Adecuacidad de los criterios de admisión a la UPR

- Trece años después de la implantación de los criterios de admisión a la UPR, se plantea la necesidad de analizar la adecuacidad de los criterios de admisión de la universidad en términos de su valor predictivo, así como para identificar posibles relaciones que puedan servir de modelos alternos.
- Frecuentemente, se utilizan los análisis cuantitativos multivariados que revelan las relaciones entre variables como un mecanismo para refinar las políticas institucionales (Toutkoushian, 2007).
- A tales efectos, se ha identificado la división recursiva (Recursive partitioning) como el mecanismo para analizar las relaciones entre las variables recopiladas a través de la Solicitud de Admisión Uniforme de la UPR y los datos académicos de escuela superior de los estudiantes que ingresan a la UPR en Cayey procedentes de escuela superior.

División Recursiva

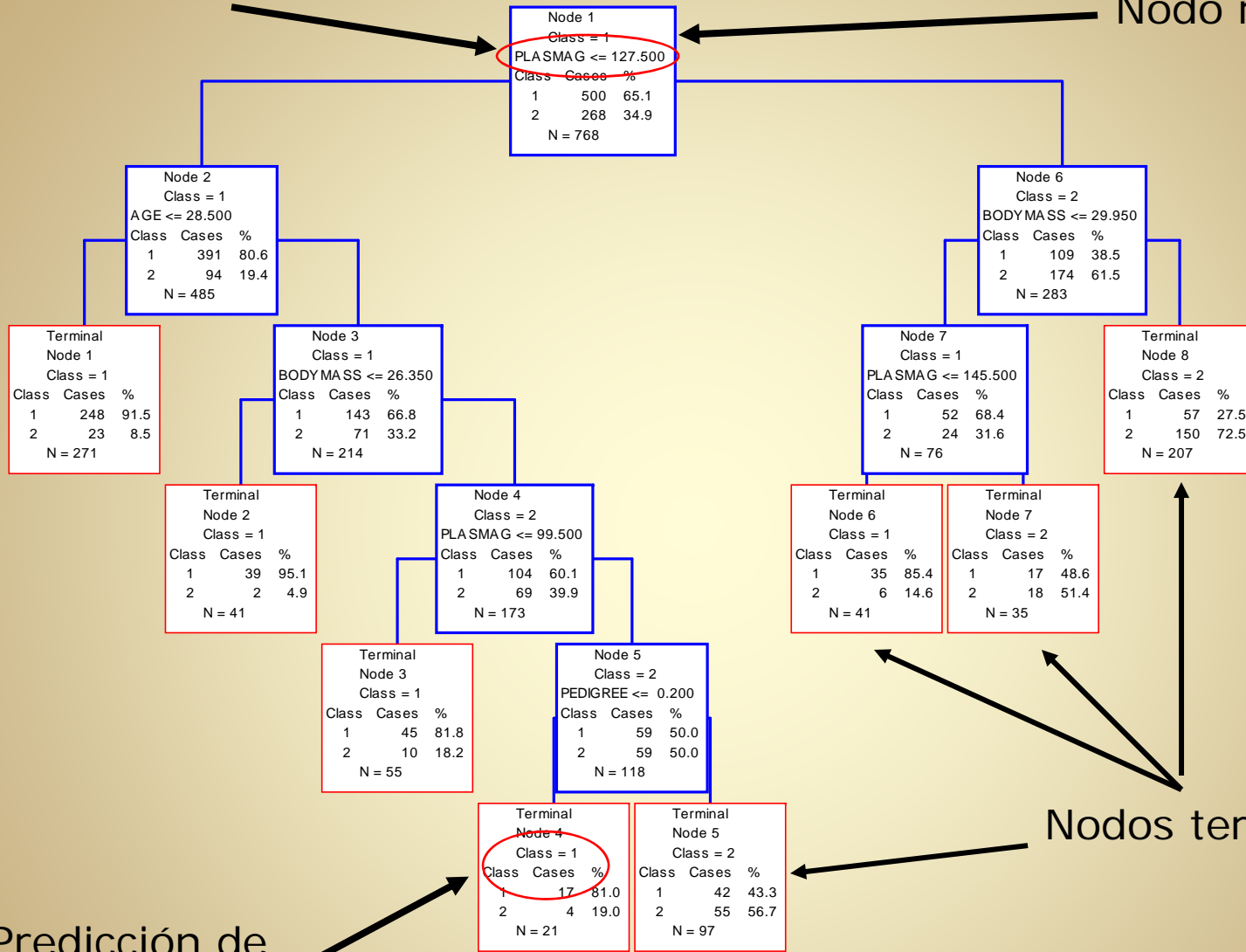
- Procedimiento mediante el cual un conjunto de datos se divide en subconjuntos cada vez más homogéneos.
- El resultado de este proceso típicamente es un dendograma, o un árbol de decisiones invertido.
- Este análisis puede utilizarse como un método alternativo al de la regresión o el análisis discriminante.

Características

- Método No-Paramétrico
- Provee flexibilidad
 - Clasificación
 - Regresión
- Las predicciones generalmente son tan buenas o mejores que las que se obtienen mediante el análisis discriminante o la regresión múltiple.
- Identifica interacciones de primer orden.

Regla de división

Nodo raíz



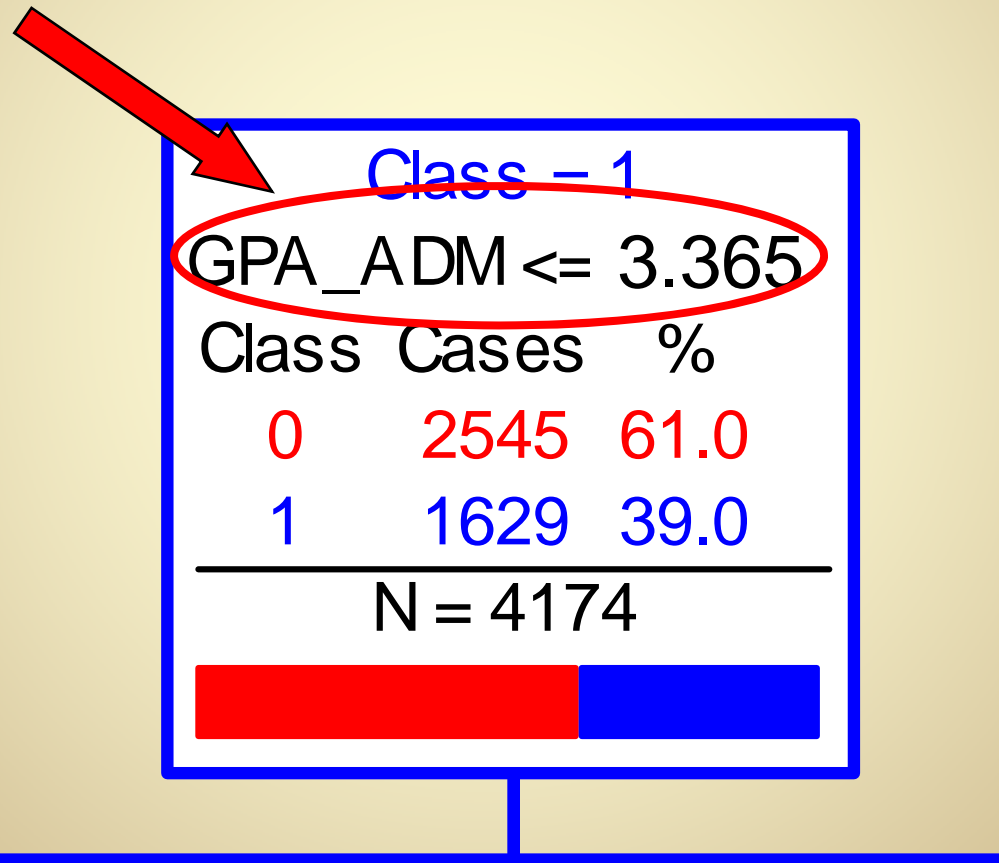
Nodos terminales

Predicción de pertenencia a clase

Árbol de Decisiones (i.e., Dendograma)

Nodo Raíz

Regla de División



Nodo Terminal

Predicción de
pertencia
a clase



Terminal
Node 4

Class	Cases	%
1	17	81.0
2	4	19.0

N = 21

Class = 1



Error de
clasificación

Cómo construir el árbol de decisiones

- Principio básico: dividir el conjunto de datos en subconjuntos homogéneos, con relación a la variable dependiente.
 - El Nodo Raíz se divide en dos sub-nodos.
 - Cada sub-nodo se divide a su vez en sub-nodos.
 - El proceso continúa hasta que algún criterio establecido se cumpla (ej., nodo es homogéneo; $n \leq 5$)
 - El árbol se poda hasta obtener el que cuente con el error de clasificación mínimo.
- El algoritmo de CART se distingue por sus divisiones binarias.

Cantidad de divisiones posibles

- El algoritmo de CART prueba todas las divisiones posibles de la variable independiente, y selecciona aquella con la “mejor” división (i.e., aquella que más reduce la heterogeneidad de los datos) .
- Para una variable continua con “n” valores distintivos, existen “n-1” divisiones posibles. Cada una de las “n-1” divisiones se realizan en el punto “ x_{jn} ”, que se encuentra en el punto medio entre dos valores consecutivos ordenados “ $x_j^{(i)}$ ” and “ $x_j^{(i+1)}$ ” .
- Para una variable discreta con “J” categorías, existen “ $2^{(J-1)} - 1$ ” divisiones posibles.

Variable Dependiente Categórica

Construcción de un árbol de decisiones mediante un ejemplo de variable dependiente categórica, con las siguientes condiciones:

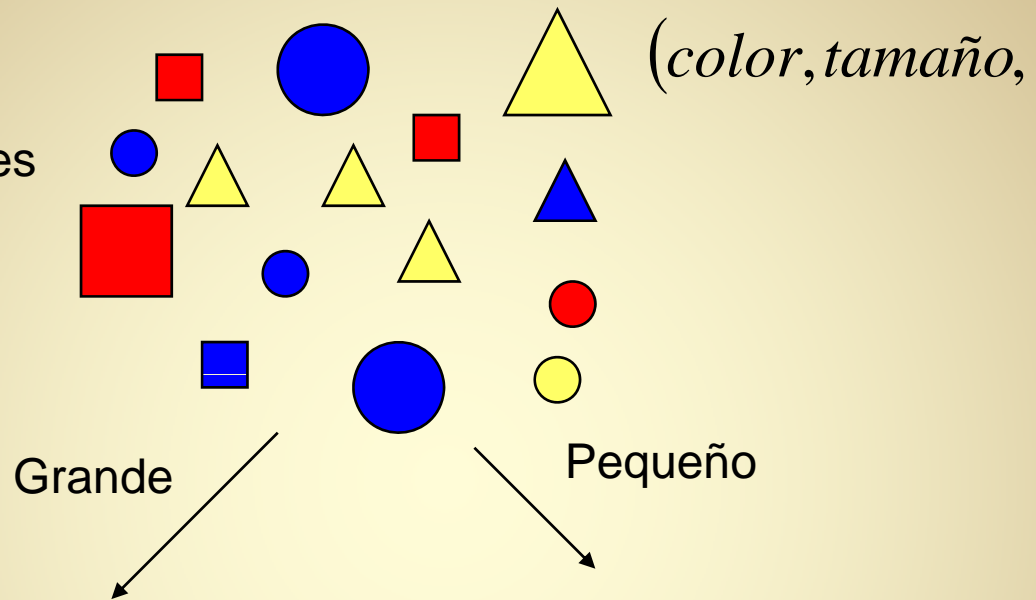
- Predicción: color (rojo, azul o amarillo)
- Predictores: tamaño (grande vs. pequeño) y figura (círculo, cuadrado o triángulo).
- $n = 15$

Tamaño

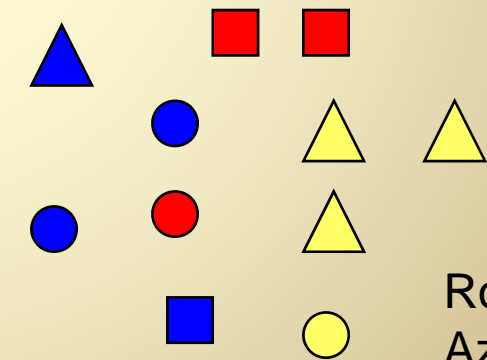
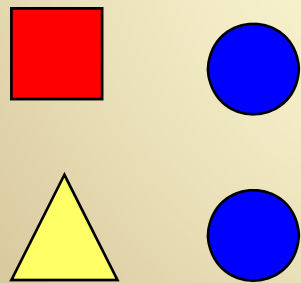
(color, tamaño, forma)

Número de Divisiones

$$2^{k-1}-1 = 2^{2-1}-1 \\ = 1$$

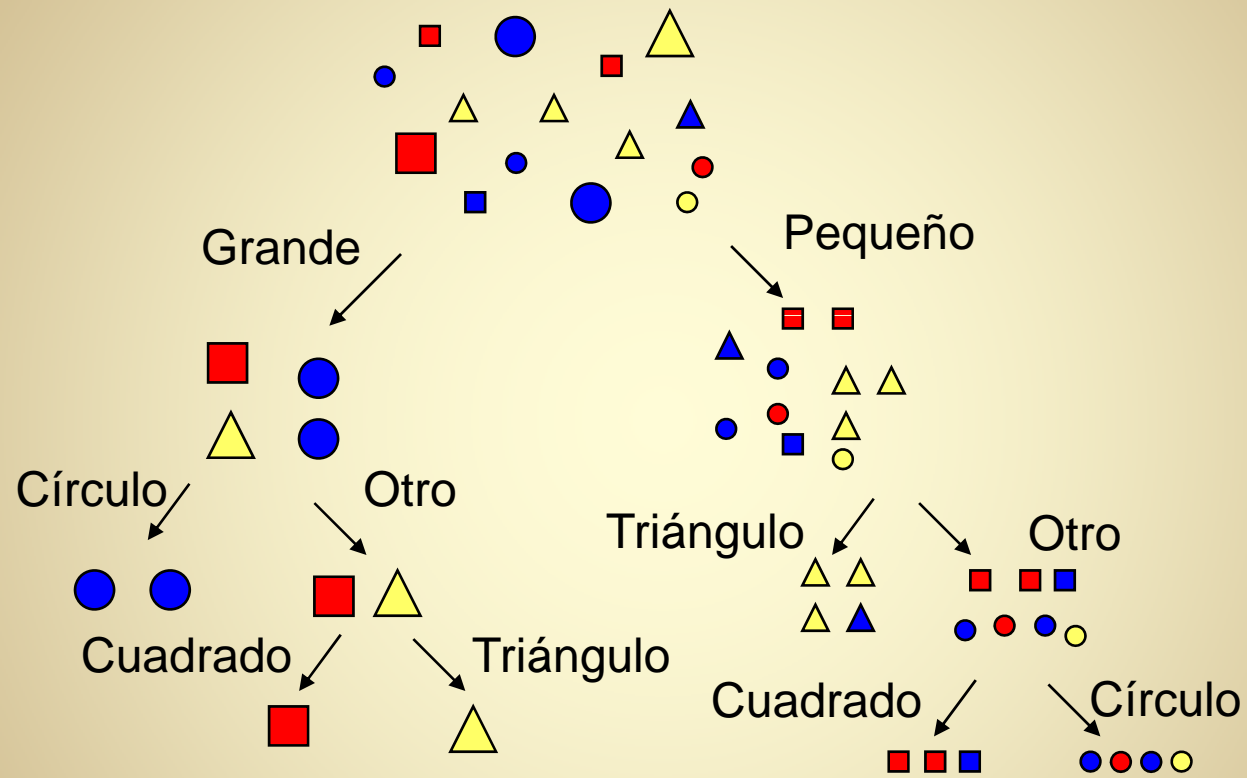


Rojo= 1
Azul= 2
Amarillo= 1



Rojo = 3
Azul= 4
Amarillo= 4

Tamaño

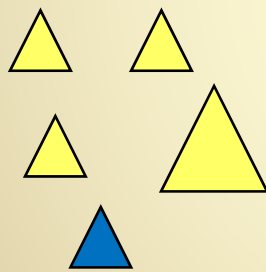


Figura, figura

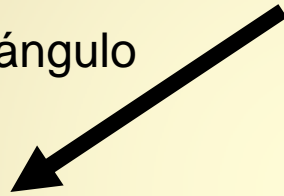
Número de Divisiones

$$2^{k-1}-1 = 2^{3-1}-1 = 3$$

Rojo= 0
Azul= 1
Amarillo= 4



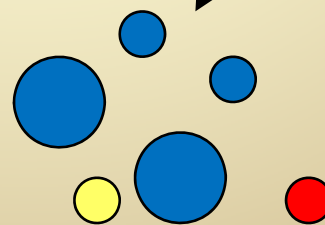
Triángulo



Otra



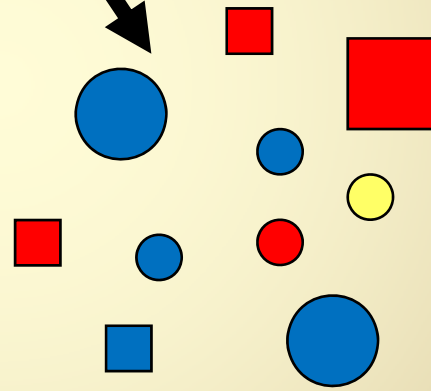
Rojo= 1
Azul= 4
Amarillo= 1



Círculo



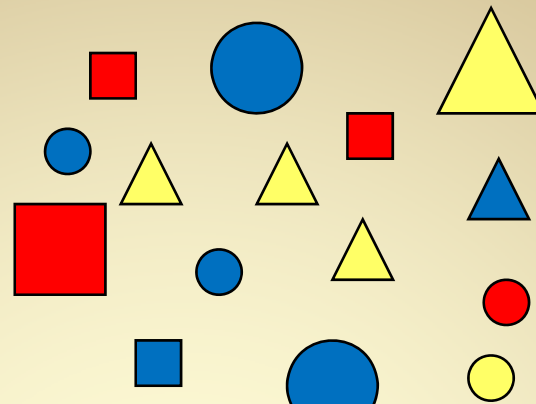
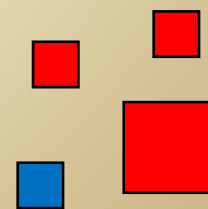
Rojo= 4
Azul= 5
Amarillo= 1



Cuadrado



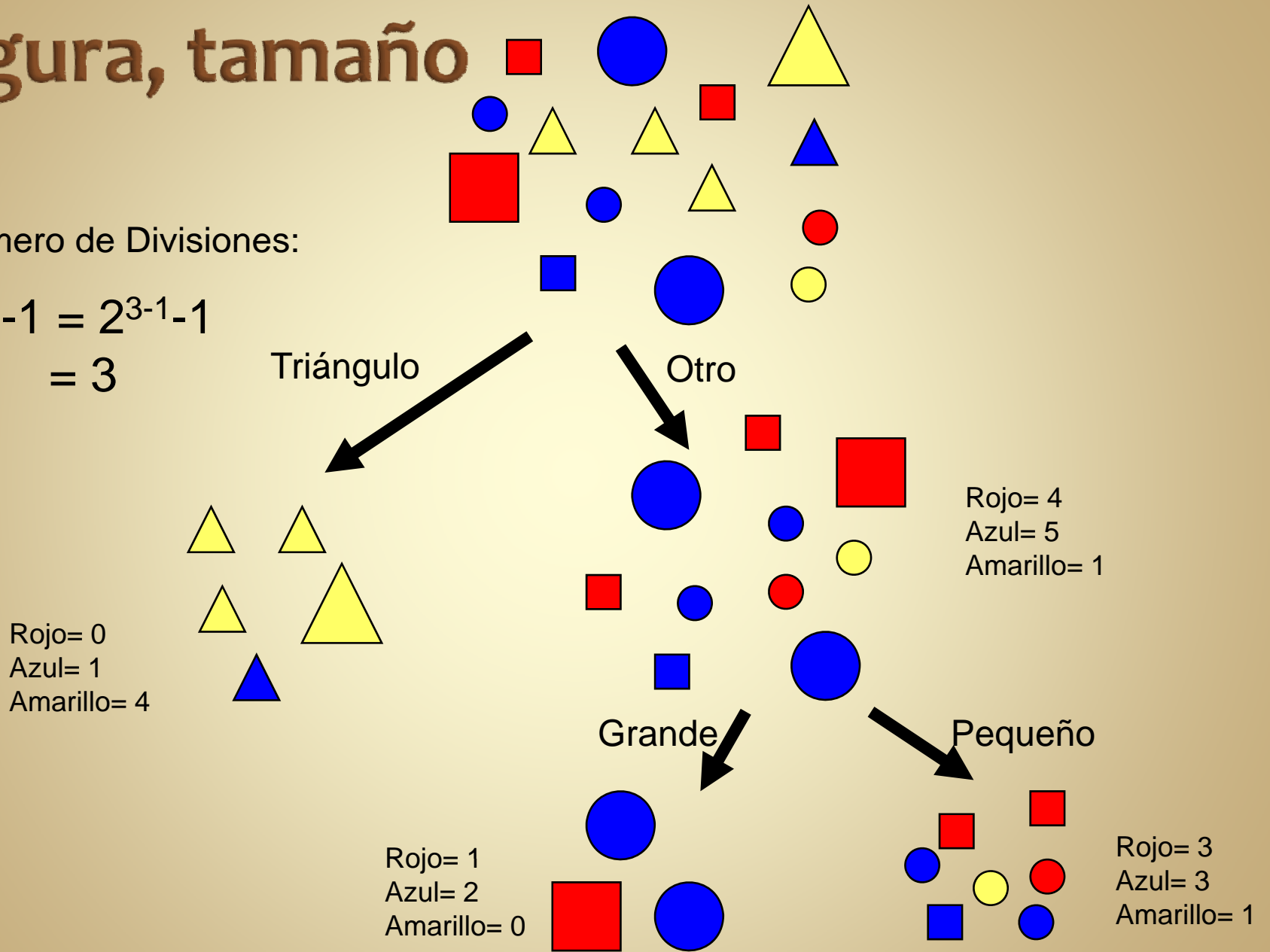
Rojo= 3
Azul= 1
Amarillo= 0



Figura, tamaño

Número de Divisiones:

$$2^{k-1}-1 = 2^{3-1}-1 = 3$$



CART

- Las divisiones ocurren por base de la reducción de la impureza de una función.
- Continúa hasta que los datos en el nodo son homogéneos, o hasta que existan muy pocos casos disponibles.
- Este árbol “máximo” es a su vez podado hasta que se obtenga un árbol con tasa de error mínima.

Árboles de Clasificación

- Evaluando la heterogeneidad de una variable dependiente categórica :
 - Se requiere una función $\Phi(p_1, p_2, \dots, p_k)$ con las siguientes características:
 - Si $p_1 = p_2 = \dots = p_k$ entonces Φ es el máximo.
 - Si $p_j = 1$ y $p_i = 0, \forall i \neq j$, entonces $\Phi = 0$.
- CART ofrece diferentes reglas de división, dos de las más utilizadas son:
 - GINI
 - Twoing

Criterios de División de GINI

- Medida de heterogeneidad (impureza) en el nodo:

$$i(t) = 1 - \sum_{j=1}^J p^2(j|t)$$

- La variable de división seleccionada es la que minimiza:

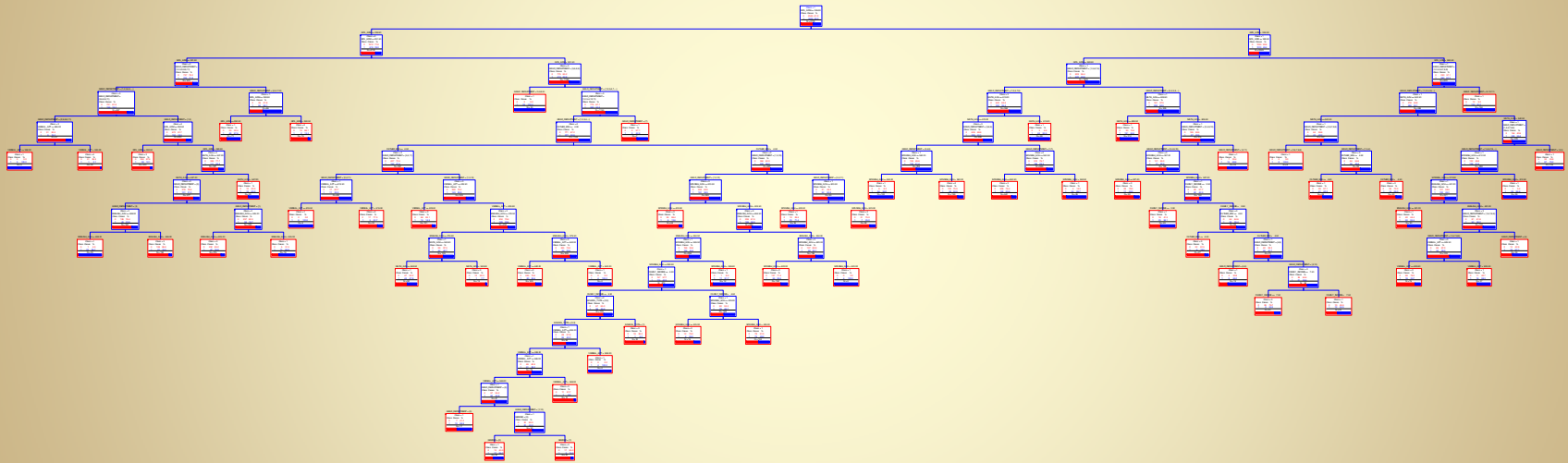
$$\Delta i = i(t) - p_1 i(1) - p_2 i(2)$$

$p(j|t)$ es la proporción de la clase “j” en el nodo “t”

Estimando el Tiempo a Completar el Grado (N = 5,240)

- Tiempo a completar grado:
 - 6.5 años o menos
 - Más de 6.5 años
- Predictores:
 - Índice de Escuela Superior
 - Puntuaciones CEEB:
 - Aptitud Matemática
 - Aptitud Verbal
 - Aprovechamiento en Inglés
 - Aprovechamiento en Español
 - Aprovechamiento Matemático
 - Educación del Padre
 - Educación de la Madre
 - Ingreso Familiar
 - Tipo de Escuela:
 - Pública
 - Privada
 - Programa de Estudio de Admisión

Árbol de Clasificación

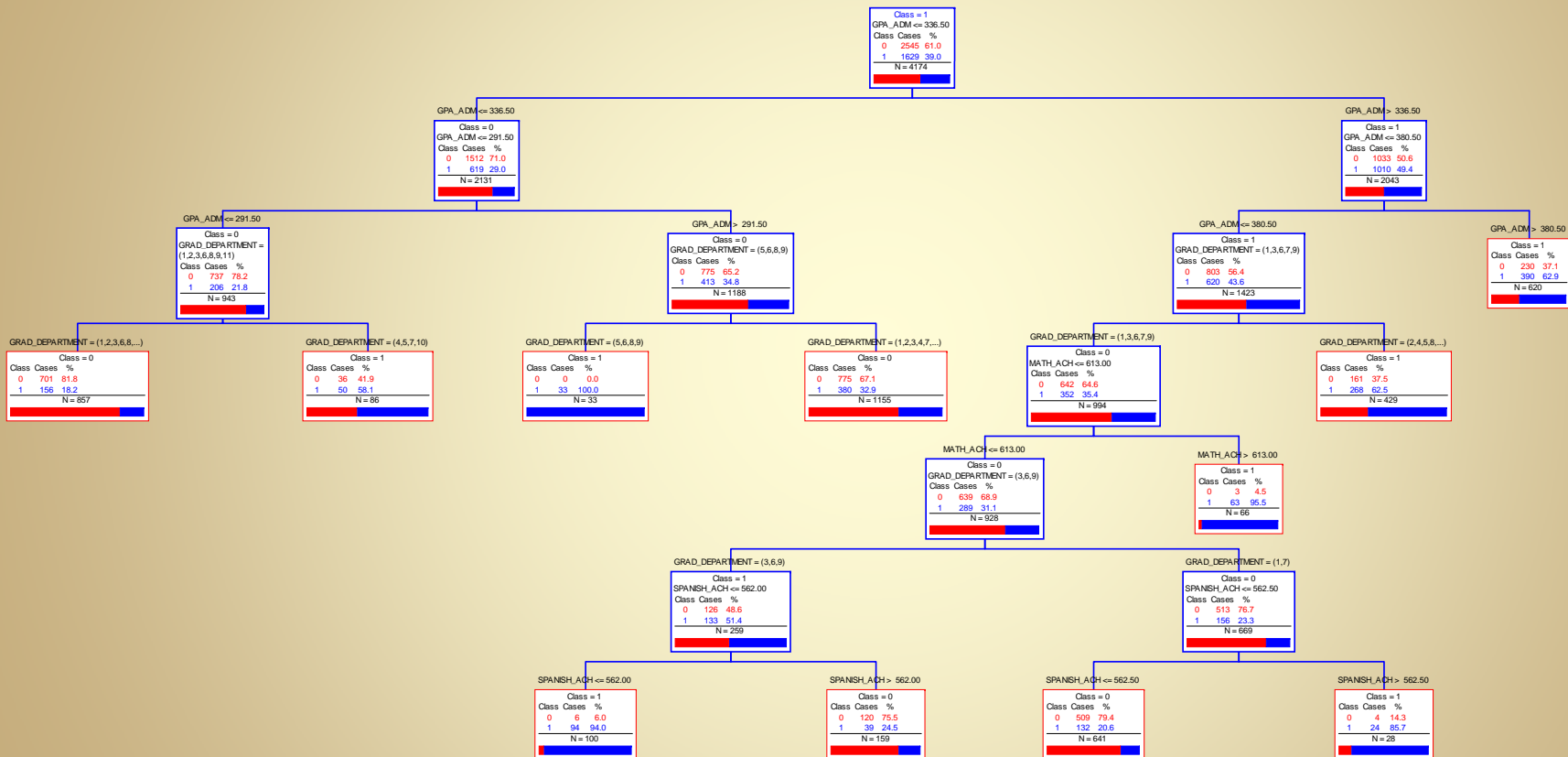


Error de Clasificación (muestra independiente) = 23.4%

Importancia de la Variable

Variable	Puntuación	
DEPARTMENT	100.00	
GPA_ADM	92.08	
SPANISH_ACH	85.81	
MATH_ACH	77.13	
VERBAL_APT	39.79	
ENGLISH_ACH	38.63	
MATH_APT	29.64	
FATHER_EDU	10.70	
FAMILY_INCOME	8.80	
SCHOOL_TYPE	5.73	
GENDER	2.57	
MOTHER_EDU	2.07	
FIRST_GENERATION	2.06	

Árbol de Clasificación

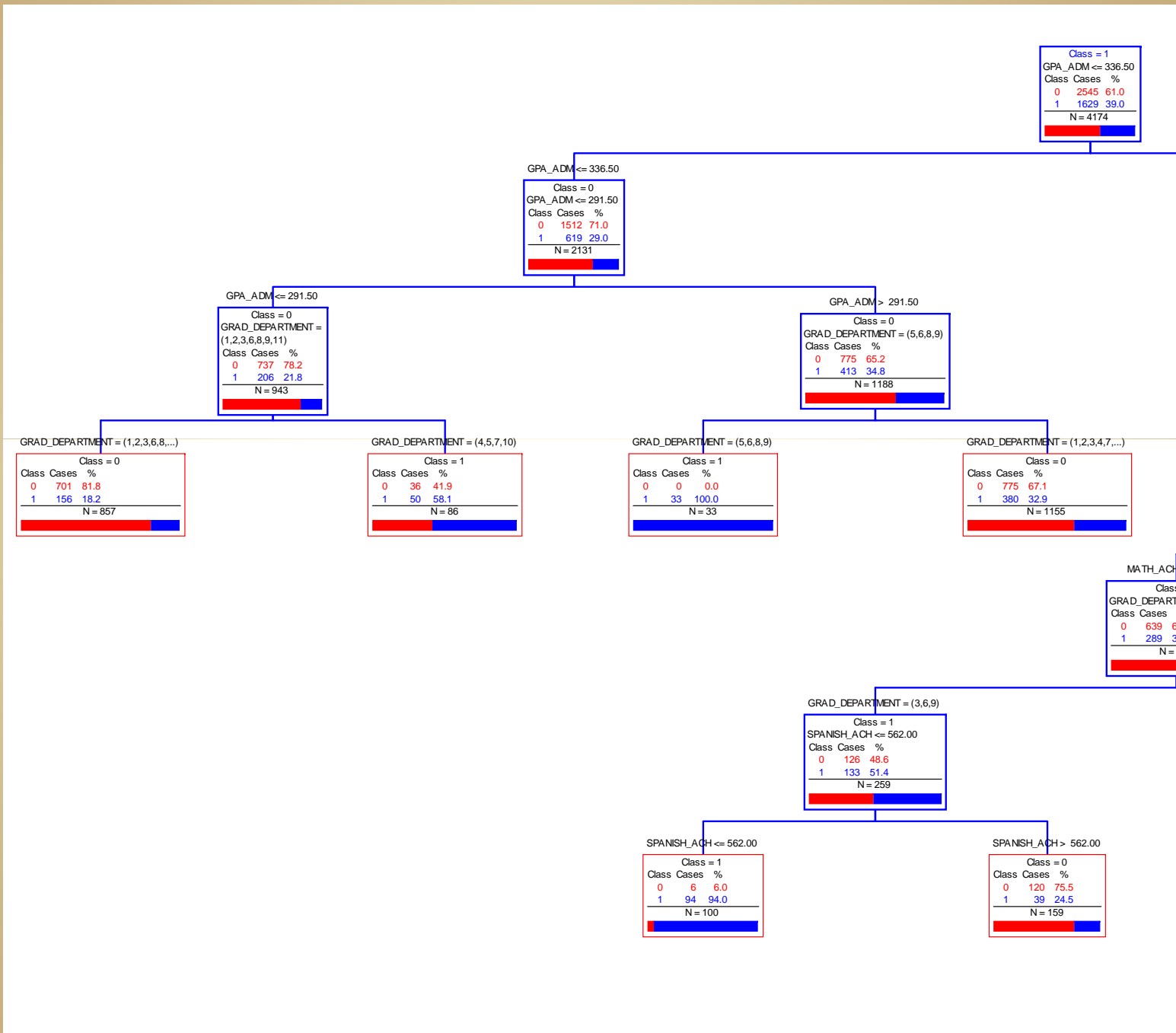


Nodo Raíz

Class = 1		
GPA_ADM <= 3.365		
Class	Cases	%
0	2545	61.0
1	1629	39.0
<hr/>		
N = 4174		

A horizontal stacked bar chart is positioned below the table. The bar is divided into two segments: a red segment on the left representing 61.0% of the total, and a blue segment on the right representing 39.0% of the total. The total length of the bar corresponds to the total number of cases, N = 4174.





Class = 1		
GPA_ADM <= 336.50		
Class	Cases	%
0	2545	61.0
1	1629	39.0
N = 4174		

Class = 1		
GPA_ADM <= 380.50		
Class	Cases	%
0	1033	50.6
1	1010	49.4
N = 2043		

Class = 1		
GRAD_DEPARTMENT = (1,3,6,7,9)		
Class	Cases	%
0	803	56.4
1	620	43.6
N = 1423		

Class = 1		
GPA_ADM > 380.50		
Class	Cases	%
0	230	37.1
1	390	62.9
N = 620		

Class = 0		
DEPARTMENT = (1,2,3,4,7,...)		
Class	Cases	%
0	775	67.1
1	380	32.9
N = 1155		

Class = 0		
MATH_ACH <= 613.00		
Class	Cases	%
0	642	64.6
1	352	35.4
N = 994		

Class = 1		
GRAD_DEPARTMENT = (2,4,5,8,...)		
Class	Cases	%
0	161	37.5
1	268	62.5
N = 429		

Class = 0		
MATH_ACH <= 613.00		
Class	Cases	%
0	639	68.9
1	289	31.1
N = 928		

Class = 1		
MATH_ACH > 613.00		
Class	Cases	%
0	3	4.5
1	63	95.5
N = 66		

Class = 0		
SPANISH_ACH <= 562.50		
Class	Cases	%
0	513	76.7
1	156	23.3
N = 669		

Class = 0		
SPANISH_ACH > 562.00		
Class	Cases	%
0	120	75.5
1	39	24.5
N = 159		

Class = 0		
SPANISH_ACH <= 562.50		
Class	Cases	%
0	509	79.4
1	132	20.6
N = 641		

Class = 1		
SPANISH_ACH > 562.50		
Class	Cases	%
0	4	14.3
1	24	85.7
N = 28		

Algunos Hallazgos

- En general, 49% de los estudiantes completan el grado dentro del 150% del tiempo esperado (i.e., 6 años).
- Si el índice de escuela superior era > 3.80 , 63% completan el grado en la UPRC.
- Si el índice de escuela superior está entre 3.37 y 3.80, 43.6% completan el grado;
 - Si además ingresaron a los departamentos de Administración de Empresas, Ciencias Sociales, Estudios Hispánicos, Inglés y Matemáticas, 62.5% completan el grado.

Algunos Hallazgos

- Si el índice de escuela superior es < 3.37 , 71% **no** completarán el grado.
- Si el índice de escuela superior está entre 2.92 and 3.36, y el departamento de admisión es:
 - Educación Física, Estudios Hispánicos, Inglés y Matemáticas, 100% ($n = 33$) completarán el grado.
 - Otro departamento, 67% ($n = 33$) no completarán el grado.
- Si el índice de escuela superior es < 2.92 , 78.2% no completarán el grado.

Valor de las Interacciones de Primer Orden

- Análisis de CART:
 - Puede utilizarse como una alternativa a la regresión, el análisis discriminante y otros métodos paramétricos para resolver problemas de predicción.
 - Puede utilizarse para complementar los análisis antes mencionados, brindando una explicación más clara de la interacción entre las variables.
- La División Recursiva nos permite obtener predicciones que son tan buenas, o mejores que aquellas arrojadas por el análisis discriminante o la regresión múltiple.

Valor de las Interacciones de Primer Orden

Las interacciones de primer orden:

- Arrojan evidencia empírica más sólida para el desarrollo de políticas institucionales respecto a criterios de admisión.
- Le ofrecen a las oficinas de Investigación Institucional y Avalúo un modelo alternativo para identificar las variables latentes y las interacciones que afectan el ambiente académico.

Valor de la División Recursiva para la Investigación Institucional

- Las interacciones de primer orden arrojan los datos necesarios para diseñar planes de intervención de calidad dirigidos a aumentar la retención y las tasas de graduación, así como para fomentar el éxito académico.
- El Análisis de CART le permite a los/las investigadores identificar diferentes escenarios de “riesgo”, redundando en la identificación y recomendación para la adjudicación de responsabilidad a aquellas unidades institucionales cuya responsabilidad es brindarle servicio a estas poblaciones.

Valor de la División Recursiva para la Administración Universitaria

- Le ofrece a la administración, la facultad, los/las consejeros/as profesionales y a otras partes interesadas (i.e., *stakeholders*) la información que sustenta el desarrollo de intervenciones, programas, servicios y políticas institucionales que fomenten el desarrollo integral de los diferentes grupos de estudiantes mediante la división recursiva.
- Provee una herramienta para que las partes interesadas (i.e., “*stakeholders*”) en los desempeños universitarios desarrollen una apreciación mayor por la importancia que tiene la investigación institucional para el mejoramiento continuo de la universidad, particularmente cuando los hallazgos de los estudios institucionales generan información compleja sobre los comportamientos estudiantiles, tales como los que se obtienen a través de la División Recursiva.

Discusión y Preguntas

Información Adicional:

Dr. José N. Caraballo, Dept. Matemáticas

jcaraballo@cayey.upr.edu

Profa. Irmannette Torres-Lugo, OAI

irmannette.torres@upr.edu

Para éste y otros datos y estudios institucionales, visite:

Oficina de Avalúo e Investigación Institucional

<http://www.cayey.upr.edu/oficina-de-avaluo-e-investigacion-institucional>

