



<Artículo>

Cómo aplicar *árboles de decisión* en SPSS.

Vanesa Berlanga Silvente, María José Rubio Hurtado, Ruth Vilà Baños

Fecha de presentación: 11/10/2012

Fecha de aceptación: 19/10/2012

Fecha de publicación: 08/01/2013

//Resumen

Un *árbol de decisión* es una forma gráfica y analítica de representar todos los eventos (sucesos) que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión más "acertada", desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Estos árboles permiten examinar los resultados y determinar visualmente cómo fluye el modelo. Los resultados visuales ayudan a buscar subgrupos específicos y relaciones que tal vez no encontraríamos con estadísticos más tradicionales.

Los *árboles de decisión* son una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas.

La función *árboles de decisión* (Tree) en SPSS crea árboles de clasificación y de decisión para identificar grupos, descubrir las relaciones entre grupos y predecir eventos futuros. Existen diferentes tipos de árbol: CHAID, CHAID exhaustivo, CRT y QUEST, según el que mejor se ajuste a nuestros datos.

//Palabras clave

Árbol de decisión, CHAID, clasificación, minería de datos.

// Referencia recomendada

Berlanga Silvente, V., Rubio Hurtado, M. J., Vilà Baños, R. (2013). Cómo aplicar árboles de decisión en SPSS. [En línea] *REIRE, Revista d'Innovació i Recerca en Educació*, 6 (1), 65-79. Accesible en: <http://www.ub.edu/ice/reire.htm>

// Datos de las autoras

Vanesa Berlanga Silvente. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). berlanga.silvente@ub.edu

María José Rubio Hurtado. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). mjrubio@ub.edu

Ruth Vilà Baños. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). ruth_vila@ub.edu

1. Dónde ubicamos los *árboles de decisión*

Los *árboles de decisión* son una técnica de *minería de datos* (*Data Mining*, DM) prepara, sondea y explora los datos para sacar la información oculta en ellos. Se aborda la solución a problemas de predicción, clasificación y segmentación.

Las técnicas de la *minería de datos* provienen de la Inteligencia Artificial y de la Estadística. Dichas técnicas no son más que algoritmos, más o menos sofisticados, que se aplican sobre un conjunto de datos para obtener unos resultados. Las técnicas más representativas son: redes neuronales, regresión lineal, *árboles de decisión*, modelos estadísticos, agrupamiento o clustering y reglas de asociación.

La clasificación inicial de las técnicas de *minería de datos* distingue entre técnicas predictivas, en las que las variables pueden clasificarse en dependientes e independientes; técnicas descriptivas, en las que todas las variables tienen el mismo estatus y técnicas auxiliares, en las que se realiza un análisis multidimensional de datos. En la figura 1 se muestra una clasificación de las técnicas de *minería de datos* donde hallamos los *árboles de decisión* (Pérez y Santín, 2008).

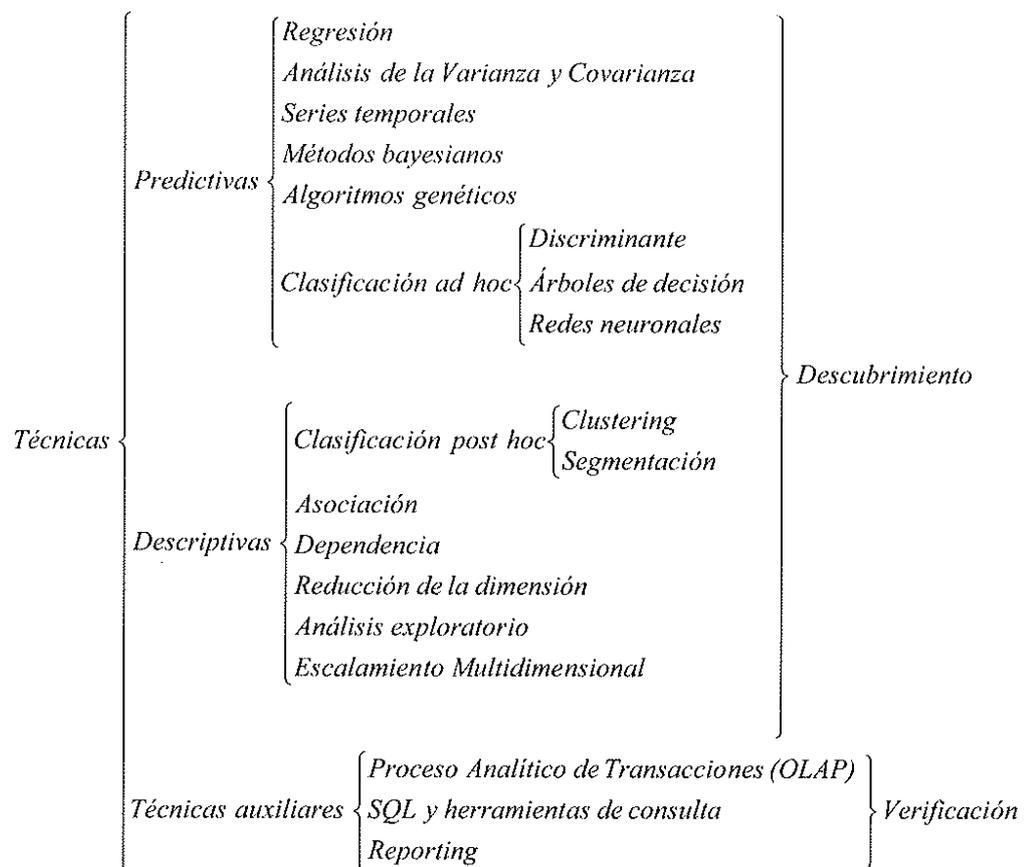


Figura 1. Clasificación de las técnicas de Data Mining (Pérez y Santín, 2008).



2. Usos generales del análisis de *árboles de decisión*

Los *árboles de decisión* crean un *modelo de clasificación* basado en diagramas de flujo. Clasifican casos en grupos o pronostican valores de una variable dependiente (criterio) basada en valores de variables independientes (predictoras).

Las ventajas de un *árbol de decisión* son (Pérez, 2011):

- Facilita la interpretación de la decisión adoptada.
- Facilita la comprensión del conocimiento utilizado en la toma de decisiones.
- Explica el comportamiento respecto a una determinada decisión.
- Reduce el número de variables independientes.

Para ejemplificar de forma gráfica lo que es un *árbol de decisión* proponemos un sencillo ejemplo. Imaginemos que queremos conocer qué variables influyen primordialmente a la hora de aprobar el primer curso de Ingeniería Electrónica. Es decir, queremos conocer qué estudiantes matriculados de primer curso de esta carrera tienen más probabilidades de aprobar el conjunto de las asignaturas, y qué características están asociadas a este éxito académico. En este caso, la variable de interés (VD) es el rendimiento académico en el primer curso. Tras introducir los datos necesarios en el programa, éste nos devuelve un diagrama que nos permite comprobar que la satisfacción con la carrera y las horas de estudio diarias son las variables que determinan principalmente el éxito académico de un estudiante. Aprueban, sobre todo, los estudiantes satisfechos con la carrera que cursan y, dentro de este grupo, los que dedican más horas al estudio. Es decir, si un estudiante de primer curso de Ingeniería Electrónica está satisfecho con la carrera y estudia más de tres horas diarias, entonces tiene una probabilidad de aprobar el conjunto del primer curso del 93%.

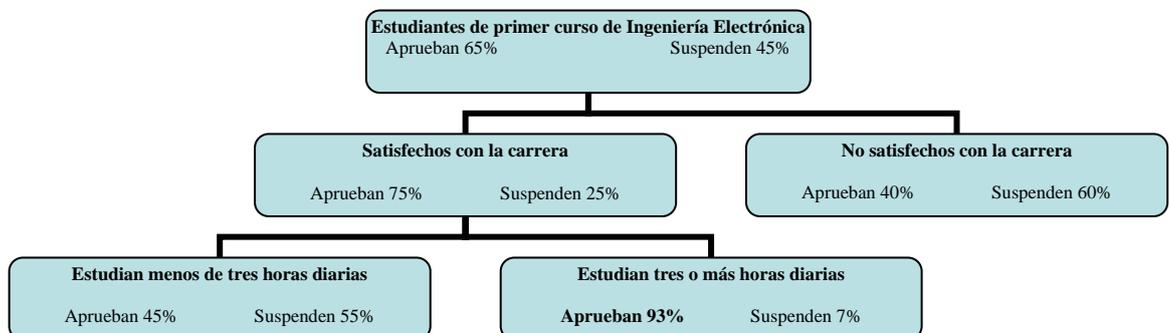


Figura 2. *Árbol de clasificación del rendimiento académico de estudiantes de primer curso de Ingeniería Electrónica (datos ficticios para uso didáctico).*



La terminología asociada a la técnica de los *árboles de decisión* recurre a una terminología específica, por lo que consideramos interesante, antes de seguir adelante, clarificarla.

- *Nodo de decisión*: Nodo que indica que una decisión necesita tomarse en ese punto del proceso. Está representado por un cuadrado.
- *Nodo de probabilidad*: Nodo que indica que en ese punto del proceso ocurre un evento aleatorio. *Probabilidades* de que ocurran los eventos posibles como resultado de las decisiones. Está representado por un círculo.
- *Nodo terminal*: Nodo en el que todos los casos tienen el mismo valor para la variable dependiente. Es un nodo homogéneo que no requiere ninguna división adicional, ya que es "puro".
- *Rama*: Nos muestra los distintos caminos que se pueden emprender cuando tomamos una decisión o bien ocurre algún evento aleatorio. *Resultados* de las posibles interacciones entre las alternativas de decisión y los eventos.

Existen cuatro métodos de división para los *árboles* cuyas características resumimos a continuación:

- *CHAID (Chi-square automatic interaction detector)*: Consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. Permite la detección automática de interacciones mediante Chi-cuadrado. En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente.
- *CHAID exhaustivo*: Supone una modificación de CHAID que examina todas las divisiones posibles para cada predictor y trata todas las variables por igual, independientemente del tipo y el número de categorías.
- *Árboles de clasificación y regresión (CRT-Classification and regression trees)*: Consiste en un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos. CRT divide los datos en segmentos para que sean lo más homogéneos posible respecto a la variable dependiente.
- *QUEST (Quick, unbiased, efficient, statistical tree)*: Consiste en un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz. Con cuatro algoritmos tenemos la posibilidad de probar métodos diferentes de crecimiento de los árboles y encontrar el que mejor se adapte a nuestros datos. Es un método rápido y que evita el sesgo que presentan otros métodos al favorecer los predictores con muchas categorías. Sólo puede especificarse QUEST si la variable dependiente es nominal.



El paquete estadístico de SPSS permite las siguientes posibilidades en relación con la técnica de los *árboles de decisión*:

- Identificar grupos, segmentos y patrones de forma altamente visual.
- Seleccionar entre diferentes tipos de árbol: CHAID, CHAID exhaustivo, CRT y QUEST, según el que mejor se ajuste a nuestros datos.
- Presentar resultados de forma intuitiva, lo que facilita la interpretación para públicos sin demasiados conocimientos de estadística.
- Guardar información de los árboles como nuevas variables en los datos. Información como el número de nodo de terminal, el valor pronosticado y las probabilidades pronosticadas.

3. Crear un *árbol de decisión*

Para acompañar la creación de un primer *árbol de decisión* y concretarlo en el programa SPSS, proponemos el siguiente caso práctico:

Se ha realizado una encuesta a 3.511 estudiantes de Ingeniería Electrónica de diversas universidades catalanas con el objetivo de identificar los factores más significativos a la hora de aprobar el primer curso de esta carrera universitaria.

Es importante que antes de empezar revisemos:

- Las escalas de medida asignadas a la matriz de datos del SPSS, ya que pueden afectar a la creación del *árbol*, si no están bien definidas.
- La muestra: se aconseja que sea suficientemente cuantiosa (evitar muestras con menos de 1.000 casos).
- La selección del método de crecimiento más adecuado: CHAID, CHAID Exhaustivo, CRT o QUEST.

Vanesa Berlanga Silvente, María José Rubio Hurtado, Ruth Vilà Baños, *Cómo aplicar árboles de decisión en SPSS*

Para ejecutar el análisis de *árbol de decisiones* se seleccionan los menús: Analizar-Clasificar-Árbol.

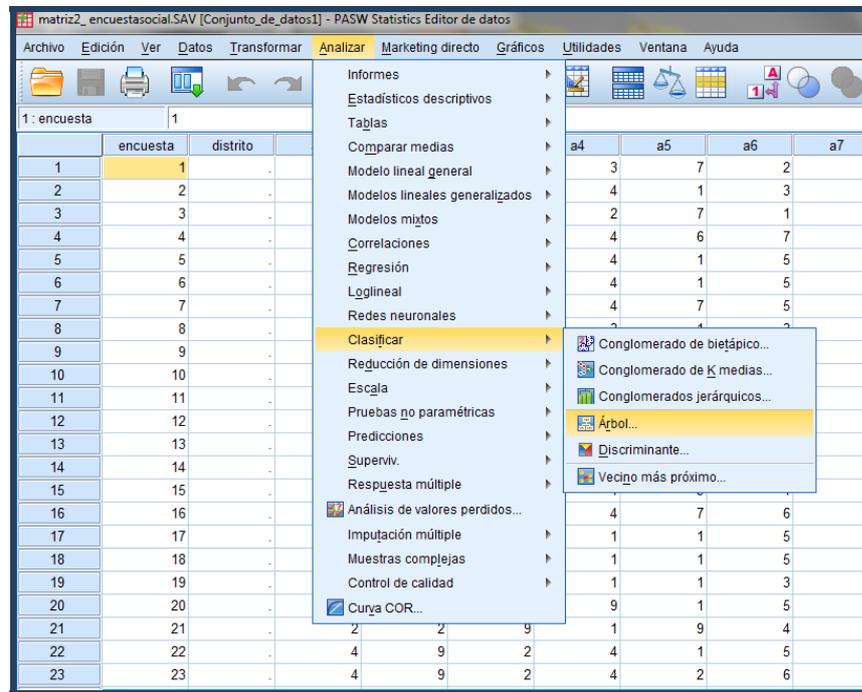


Figura 3. Cuadro de diálogo árbol de decisión en SPSS.

La variable que queremos explicar, es decir, la *variable dependiente* es el *rendimiento académico de los estudiantes de primer curso de Ingeniería Electrónica*, y como variables explicativas, esto es, las variables independientes, seleccionamos algunas del total de variables de tipo sociodemográfico incluidas en la encuesta, las consideradas más probables. En nuestro ejemplo seleccionamos las siguientes variables: *número de horas de estudio diarias, elección de la carrera en primera opción, trabajar y satisfacción con la carrera*.



Figura 4. Cuadro de diálogo árbol de decisión CHAID exhaustivo.



El procedimiento excluirá de forma automática cualquier variable de las seleccionadas cuya contribución al modelo final no sea significativa. En este momento ya se puede ejecutar el procedimiento y generar un modelo de árbol básico, pero vamos a modificar algunos parámetros que podemos seleccionar en los diferentes botones del cuadro de diálogo general de la figura 4.

En el botón *Categorías* se pueden especificar las categorías objetivo de interés de la variable dependiente. Hay que tener en cuenta que, si bien las categorías objetivo no afectan al modelo del árbol propiamente dicho, algunos resultados y opciones sólo estarán disponibles si se han seleccionado categorías objetivo. Es decir, para variables dependientes categóricas (nominales, ordinales), se puede controlar qué categorías se incluirán en el análisis e identificar las categorías objeto de interés.

La opción *primera variable forzosa* sirve para forzar que la primera variable independiente de la lista sea la primera que aparezca en la división del árbol. Esta opción únicamente se marcará cuando la investigación así lo requiera.

El desplegable *método de crecimiento* permite seleccionar entre los cuatro existentes: CHAID, CHAID exhaustivo, CRT y QUEST. Para este ejemplo, utilizaremos el método *CHAID exhaustivo*.

Resultados

Pulsando el botón de *Resultados* (en la figura 4 marcado con "a") se abre un cuadro de diálogo con pestañas, en el que se pueden seleccionar distintos tipos de opciones.

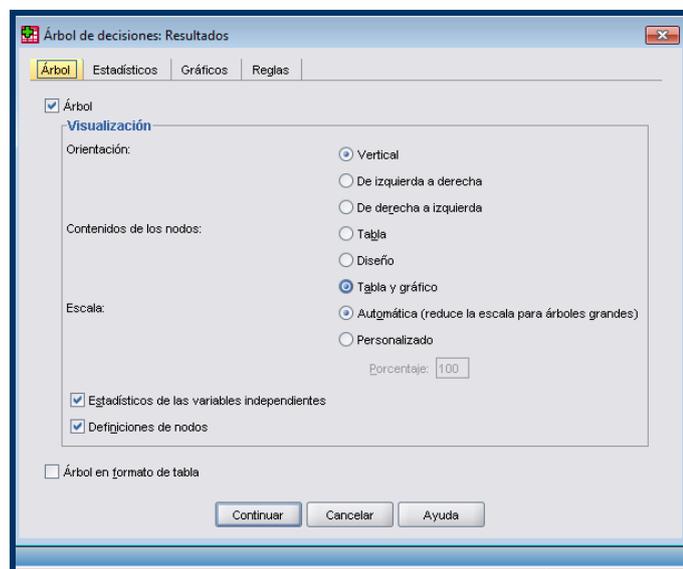


Figura 5. Cuadro de diálogo Resultados, pestaña árbol.



La pestaña *árbol* permite controlar el aspecto inicial del árbol o suprimir completamente su presentación. Por defecto aparecen ya marcadas las siguientes opciones:

- *Orientación*. El árbol se muestra de arriba abajo, con el nodo raíz situado en la parte superior. También se podría mostrar de izquierda a derecha, o de derecha a izquierda.
- *Contenidos de los nodos*. Los nodos pueden mostrar tablas, gráficos o ambos. Para variables dependientes categóricas, las tablas muestran frecuencias y porcentajes, y los gráficos son diagramas de barras. Para variables dependientes de escala, las tablas muestran medias, desviaciones típicas, número de casos y valores pronosticados, y los gráficos son histogramas. Por defecto, aparece la tabla, aunque sugerimos seleccionar *árbol en formato de tabla y gráfico*. Esta opción ofrece una opción gráfica muy clarificadora del árbol.
- *Escala*. Por defecto, los árboles grandes se reducen de forma automática para intentar ajustar el árbol a la página, pero se puede especificar un porcentaje de escala personalizado de hasta el 200%.
- *Estadísticos de las variables independientes*. Para CHAID y CHAID exhaustivo, los estadísticos incluyen el valor F (para variables dependientes de escala) o el valor Chi-cuadrado (para variables dependientes categóricas) así como el valor de significación y los grados de libertad. Para CRT, se muestra el valor de mejora. Para QUEST, se muestra el valor F, el valor de significación y los grados de libertad para las variables independientes ordinales y de escala; por su parte, para las variables independientes nominales se muestra el valor Chi-cuadrado, el valor de significación y los grados de libertad.
- *Definiciones de los nodos*. Las definiciones de nodos muestran el valor o valores de la variable independiente utilizados en cada división de nodos.

En la pestaña *Estadísticos* las opciones disponibles dependen del nivel de medida de la variable dependiente, del método de crecimiento y de otros valores de configuración. Por defecto están seleccionadas las siguientes opciones:

- *Resumen*. El resumen incluye el método utilizado, las variables incluidas en el modelo y las variables especificadas pero no incluidas en el modelo.
- *Riesgo*. Estimación del riesgo y su error típico. Una medida de la precisión predictiva del árbol. Para variables dependientes categóricas, la estimación de riesgo es la proporción de casos clasificados incorrectamente después de haber sido corregido respecto a las probabilidades previas y los costes de clasificación errónea. Para variables dependientes de escala, la estimación de riesgo corresponde a la varianza dentro del nodo.
- *Tabla de clasificación*. Para variables dependientes categóricas (nominales, ordinales), esta tabla muestra el número de casos clasificados correcta e



Vanesa Berlanga Silvente, María José Rubio Hurtado, Ruth Vilà Baños, *Cómo aplicar árboles de decisión en SPSS*

incorrectamente para cada categoría de la variable dependiente. No está disponible para variables dependientes escalares.

La pestaña *Reglas* ofrece la capacidad de generar reglas de selección o clasificación/predicción en forma de sintaxis de comandos, SQL o sólo texto (inglés sin formato). Estas reglas se pueden visualizar en el Visor y/o guardar en un archivo externo.

Validación del modelo

En el botón *Validación* (en la figura 4 marcado con "b") podemos validar el modelo. La validación permite evaluar la bondad de la estructura de *árbol* cuando se generaliza para una mayor población. Existen dos métodos de validación disponibles: la validación cruzada y la validación por división muestral.

La *validación cruzada* divide la muestra en un número de submuestras y, a continuación, se generan los modelos de árbol.

Con la *validación por división muestral*, el modelo se genera utilizando una muestra de entrenamiento y después pone a prueba ese modelo con una muestra de reserva.



Figura 6. Cuadro de diálogo validación.

Criterios de crecimiento del árbol

El botón *Criterios* (en la figura 4 marcado con "c") permite establecer los criterios de crecimiento del árbol. Para este ejemplo, deseamos que el árbol sea lo más sencillo posible, así que limitaremos el crecimiento del árbol elevando el número de casos mínimo para nodos parentales y filiales, tal como aparece por defecto en el programa.



La pestaña *Límites de crecimiento* permite limitar el número de niveles del árbol y controlar el número de casos mínimo para nodos parentales y filiales. Se pueden modificar algunos de los siguientes parámetros:

- *Máxima profundidad de árbol.* Controla el número máximo de niveles de crecimiento por debajo del nodo raíz. El ajuste *Automática* limita el árbol a tres niveles por debajo del nodo raíz para los métodos CHAID y CHAID exhaustivo y a cinco niveles para los métodos CRT y QUEST.
- *Número de casos mínimo.* Controla el número de casos mínimo para los nodos. Los nodos que no cumplen estos criterios no se dividen. El aumento de los valores mínimos tiende a generar árboles con menos nodos. La disminución de dichos valores mínimos generará árboles con más nodos. Para archivos de datos con un número pequeño de casos, es posible que, en ocasiones, los valores por defecto de 100 casos para nodos parentales y de 50 casos para nodos filiales den como resultado árboles sin ningún nodo por debajo del nodo raíz; en este caso, la disminución de los valores mínimos podría generar resultados más útiles.

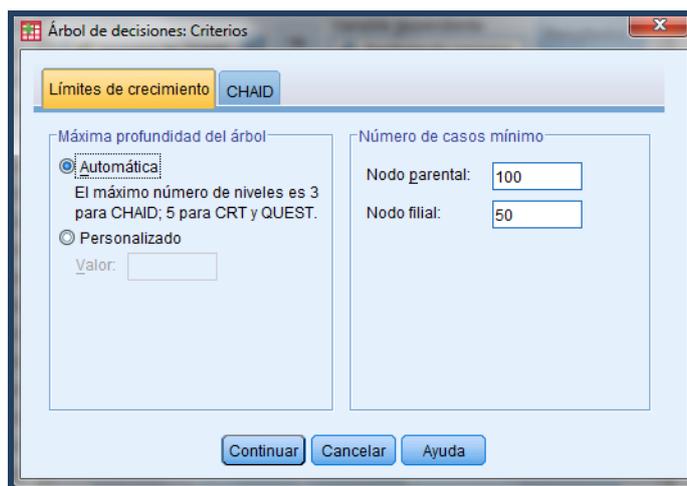


Figura 7. Cuadro de diálogo Criterios, pestaña Límites de crecimiento.

Para los métodos CHAID y CHAID exhaustivo (pestaña CHAID), puede controlarse el nivel de significación para la división de nodos y la fusión de categorías. Para ambos criterios, el nivel de significación por defecto es igual a 0,05.

Debe tenerse en cuenta que, para variables dependientes ordinales, el valor de Chi-cuadrado para determinar la división de nodos y la fusión de categorías se calcula mediante el método de la razón de verosimilitud. Para variables dependientes nominales, puede seleccionarse el método Pearson.



4. Cómo interpretar los *outputs*

Al generar el árbol obtenemos tres outputs en SPSS:

- Tabla que proporciona información acerca del modelo.
- Diagrama del árbol.
- Las variables de predicción del modelo añadidas al conjunto de datos activo.

La *Tabla de resumen* del modelo proporciona información general sobre las especificaciones utilizadas para crear el modelo y sobre el modelo resultante. La sección *Especificaciones* ofrece información sobre los valores de configuración utilizados para generar el modelo de *árbol*, incluidas las variables utilizadas en el análisis. La sección *Resultados* muestra información sobre el número de nodos totales y terminales, la profundidad del *árbol* (número de niveles por debajo del nodo raíz) y las variables independientes incluidas en el modelo final.

Resumen del modelo

Especificaciones	Método de crecimiento	CHAID exhaustivo	
	Variable dependiente	Rendimiento académico primer curso	
	Variables independientes	Trabaja, Satisfacción con la carrera, Elección de la carrera en primera opción, N° de horas diarias de estudio	
	Validación	Ninguna	
Resultados	Máxima profundidad de árbol		3
	Mínimo de casos en un nodo filial		100
	Mínimo de casos en un nodo parental		50
	Variables independientes incluidas	Satisfacción con la carrera, N° de horas diarias de estudio, Elección de la carrera en primera opción, Trabaja	
	Número de nodos		13
	Número de nodos terminales		7
	Profundidad		3

Figura 8. Tabla Resumen del modelo..

El Diagrama de *árbol* obtenido es una representación gráfica del modelo del árbol. En el ejemplo, todas las variables son tratadas como nominales y cada nodo contiene una tabla de frecuencias que muestra el número de casos (frecuencia y porcentaje) para cada categoría de la variable dependiente. También incluye el gráfico de frecuencias.

La categoría "pronosticada", que es la categoría con el mayor valor de frecuencia en cada nodo, aparece resaltada con una franja gris.



Vanesa Berlanga Silvente, María José Rubio Hurtado, Ruth Vilà Baños, *Cómo aplicar árboles de decisión en SPSS*

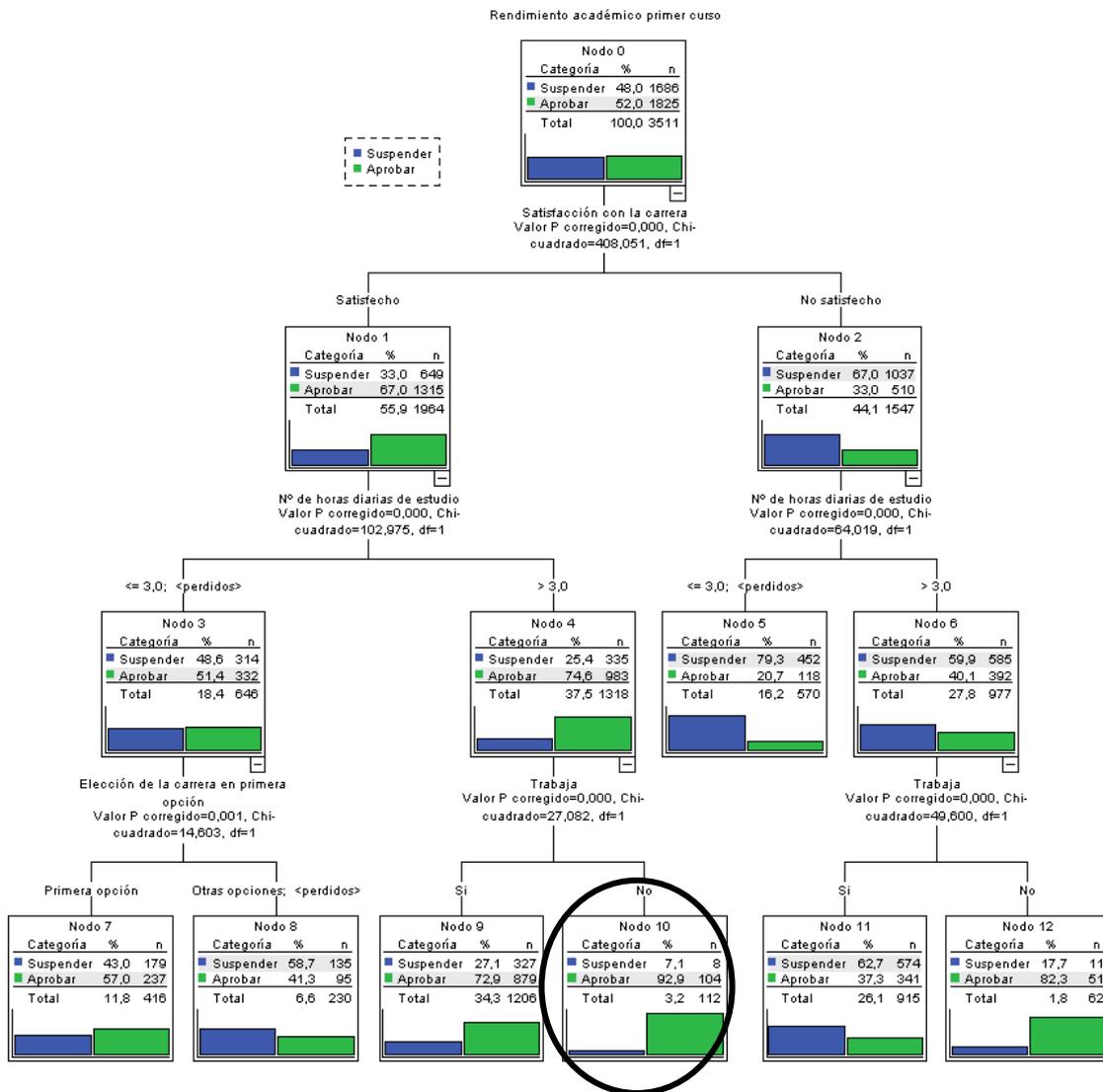


Figura 9. Diagrama del árbol para el modelo de éxito académico en el primer curso de Ingeniería Electrónica.

Cómo empezar a interpretar el árbol de decisión:

1. En primer lugar, nos fijamos en el nodo 0 que describe la variable dependiente: porcentaje de los estudiantes que suspenden y de los que aprueban.
2. Seguidamente observamos que la variable dependiente se ramifica en dos nodos: Nodo 1 y 2 pertenecientes a la variable *Satisfacción con la carrera*, indicando que ésta es la variable principal predictora.



3. A continuación, debemos fijarnos en el Nodo 1, ya que su Chi-Cuadrado es superior a la del Nodo 2. Además, nos interesa conocer el perfil de los estudiantes que aprueban, por ser nuestro objetivo de investigación. El Nodo 1 nos indica que del 55.9% de los que están satisfechos con su carrera, el 67% aprueban.
4. El Nodo 1 se vuelve a ramificar en los Nodos 3 y 4 pertenecientes a la variable *Horas de estudio*. Observamos en el Nodo 4 que aprueban más los estudiantes que dedican más de tres horas diarias al estudio, con un 74.6%, frente a un 51.4% del Nodo 3 que aprueban estudiando menos de tres horas diarias.
5. El Nodo 4 se ramifica en los nodos 9 y 10, pertenecientes a la variable *Si trabaja o no*. Y aquí observamos que un 92.9% de los estudiantes que no trabajan aprueban.
6. Por tanto, a modo resumen, los nodos que definen el perfil de los estudiantes que aprueban (variables que influyen en Aprobar) son: Nodo 0 -Nodo 1 - Nodo 4 - Nodo 10. Es decir, influyen las siguientes variables: Rendimiento académico - Satisfacción con la carrera - Número de horas de estudio - Si trabaja o no.

Algunas conclusiones posibles del *árbol* de la figura 10, son:

- La variable *Satisfacción* con la carrera es el mejor predictor para *el Rendimiento académico del primer curso*, con dos categorías: *aprobar* y *suspender*.
- La probabilidad más alta de *aprobar* (92.9%) se da entre los estudiantes que se sienten satisfechos con la carrera, estudian más de tres horas diarias y no trabajan (tal como hemos señalado en el círculo de la figura 9).
- La probabilidad más baja de *aprobar* (20.7%) se da entre los estudiantes que no están satisfechos con la carrera y que dedican menos de tres horas diarias al estudio. Si estos alumnos estudian más de tres horas diarias y además no trabajan, la probabilidad de aprobar aumenta hasta el 82.3%.
- Entre los estudiantes que están satisfechos con la carrera, los que estudian menos de tres horas y no han escogido la carrera como primera opción sólo tienen el 41.3% de probabilidad de aprobar.

Para finalizar, las *variables de predicción del modelo* aparecen en las tablas de riesgo y de clasificación, y proporcionan una rápida evaluación de la bondad del funcionamiento del modelo, tal y como se observa en la figura 10.



Vanesa Berlanga Silvente, María José Rubio Hurtado, Ruth Vilà Baños, *Cómo aplicar árboles de decisión en SPSS*

Riesgo

Método	Estimación	Típ. Error
Resustitución	,307	,008
Validación cruzada	,308	,008

Métodos de crecimiento: CHAID exhaustivo
Variable dependiente: Rendimiento académico primer curso

Clasificación

Observado	Pronosticado		
	Suspender	Aprobar	Porcentaje correcto
Suspender	1161	525	68,9%
Aprobar	554	1271	69,6%
Porcentaje global	48,8%	51,2%	69,3%

Métodos de crecimiento: CHAID exhaustivo
Variable dependiente: Rendimiento académico primer curso

Figura 10. Tablas de riesgo y clasificación del árbol de decisión.

Para la interpretación de estas tablas debemos considerar que los resultados en la tabla de clasificación son coherentes con la estimación de riesgo. La tabla muestra que *el modelo clasifica de forma correcta, aproximadamente, al 69.3% de los individuos en general*. De forma específica para cada categoría de la variable dependiente ofrece un "acierto" ligeramente más elevado en el caso de la categoría "aprobar", con un 69.6%.

<Referencias bibliográficas>

- Hernández, J.; Ramírez, M.J. y Ferri, C. (2004). *Introducción a la minería de datos*. Madrid: Pearson educación.
- Lind, D.A; Marchal, W.G. y Wathen, S.A. (2012). *Estadística aplicada a los negocios y la economía*. México D.F.: MCGraw Hill.
- Pérez, C. (2011). *Técnicas de segmentación. Conceptos, herramientas y aplicaciones*. Madrid: Gaceta Grupo Editorial.
- Pérez, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid: Pearson educación.
- Pérez, C. y Santín, D. (2007). *Minería de Datos: Técnicas y Herramientas*. Madrid: Ediciones Paraninfo, S.A.



Vanesa Berlanga Silvente, María José Rubio Hurtado, Ruth Vilà Baños, *Cómo aplicar árboles de decisión en SPSS*

Silberschatz, A. (2007). *Fundamentos de diseño de bases de datos* (5ª ed.). Madrid: Mcgraw-hill/ Interamericana de España, S.A.

Torrado, M. (2011). *Minería de datos aplicados a la educación*. Consultado el 1 de octubre de 2012, en Depósito digital de la UB <http://hdl.handle.net/2445/19862>

Valderrey, P. (2010). *Técnicas de segmentación de mercados*. Madrid: Starbook editorial.

Vilà, R. y Bisquerra, R. (2004). El análisis cuantitativo de los datos. Bisquerra, R. (coord). *Metodología de la investigación educativa* (pp. 259-274). Madrid: La Muralla.

Vilà, R. (2012). *Arbres de decisió amb SPSS*. Consultado el 1 de octubre de 2012, en Depósito digital de la UB: <http://hdl.handle.net/2445/22282>.

Copyright © 2013. Esta obra está sujeta a una licencia de Creative Commons mediante la cual, cualquier explotación de ésta, deberá reconocer a sus autores, citados en la referencia recomendada que aparece al inicio de este documento.

