

ARBOLES DE CLASIFICACION CON FACTOR DE PONDERACION APLICADO AL ESTUDIO DEL CONSUMO DE TABACO EN JOVENES DE LA REGION METROPOLITANA, CHILE*

L.I. Schiattino; R.R. Villegas y N.L. Caris

Escuela de Salud Pública, Facultad de Medicina, Universidad de Chile

RESUMEN

El objetivo del trabajo es describir el perfil de los estudiantes de la Región Metropolitana (R.M.) que han fumado cigarrillo o han consumido algún producto tabaco en el último mes, utilizando árboles de clasificación con factor de ponderación, sobre la base de datos de la Encuesta de Tabaquismo en Jóvenes 2000 (EMTAJOVEN, Marzo 2000, OMS, MINSAL). La muestra fue de 3150 estudiantes de 12 a 15 años. Además del factor de ponderación se utilizaron 26 variables categóricas, relacionadas a características personales y del consumo de tabaco. Se especificaron iguales probabilidades a priori y los costos de mala clasificación fueron 1, 1.5, 2, 2.5 y 3.0. Para la validación se utilizó una muestra de prueba y se documentaron las tasas de especificidad, sensibilidad y clasificación correcta sobre las muestras expandidas de construcción y validación de 13 árboles diferentes. El árbol de clasificación final determinó ocho variables para describir distintos grupos de estudiantes con una tasa de especificidad de 80% y una sensibilidad de 89%. Las variables importantes están relacionadas a características personales, lugar donde fuma, intención de fumar, exposición al humo de tabaco, creer que el cigarrillo light es menos malo y que los adolescentes que fuman tienen más o menos amigos.

ABSTRACT

The objective of this paper is to describe the profile of the Region Metropolitana's (RM) students that have smoked cigarettes or that have consumed some kind of tobacco during the last month. Classification Trees with Weight factor are used, using the database of the Tobacco Consumption Inquiry by Youngs 2000 (EMTAJOVEN, March 2000, OMS, MINSAL). The sample consisted in 3150 students between 12 and 15 years old. 26 categorical variables, related with personal characteristics and the consumption of tobacco were measured and the weight factor. Equal a priori probabilities were specified and the incorrect classification costs were 1, 1.5, 2, 2.5 ad 3.0. For validating, a test sampling was used and the specificity, sensitivity and correct classification rates were tested on the construction and validation of 13 different trees the expanded samples. The final classification tree determined 8 variables for describing the different student groups with a specificity rate of 80% and a sensitivity of 89%. The important variables are related with personal characteristics, place where he/she smokes, intention of smoking, exposition to tobacco's smoke, the believe that to smoke light cigarettes is less dangerous and that smoker adolescents have more or less friends.

Key words: classification trees, testing, classification rates.

MSC: 62P10

1. INTRODUCCION

Estudios recientes muestran que alcohol y tabaco son las sustancias más consumidas por los escolares, presentando variaciones a nivel nacional y en la Región Metropolitana. Esto se observa si se considera el consumo de alguna vez en la vida, el consumo en el último año y en el último mes¹. Además, el 54 % de los adolescentes fumaría, poniendo a Chile a la cabeza de América Latina².

La Encuesta sobre tabaquismo en jóvenes 2000 (EMTAJOVEN, Marzo 2000, OMS, MINSAL) fue aplicada en la Región Metropolitana (R.M.) con el fin de evaluar la evolución del consumo de tabaco en los jóvenes, revelando que el 71.1 % de los estudiantes han fumado alguna vez cigarrillos, el 7.4 % ha consumido otro producto tabaco, 61.3 % vive en un medio ambiente contaminado con humo de cigarrillos, 71.9 % está rodeado de otros que fuman cigarrillo en lugares diferentes a su hogar y 70.7 % piensa que se debiera prohibir fumar cigarrillo en lugares públicos, entre otras características.

El objetivo del presente trabajo es describir el perfil de los estudiantes de la Región Metropolitana, que han fumado cigarrillo o consumido algún producto tabaco en el último mes, utilizando el método de árboles de clasificación con factor de ponderación sobre la base de datos de EMTAJOVEN.

*This paper was presented at VIII Conferencia Latinoamericana de Probabilidades y Estadística Matemática.

2. ARBOLES DE CLASIFICACION Y REGRESION, CART

CART es un procedimiento estadístico introducido por Leo Brieman, Jerome Friedman, Richard Olshen y Charles Stone en 1984.

CART es un análisis no paramétrico de clasificación binaria que puede ser utilizado con variables independientes categóricas y / o continuas. Nos referimos a árboles de regresión si la variable respuesta es continua y árboles de clasificación si la variable respuesta es categórica. Se caracteriza por no requerir selección a priori de variables, presencia de interacciones o transformación de variables. Es robusto frente a los valores extremos, errores de medición y presencia de variables predictoras irrelevantes. Puede usar la misma variable en diferentes partes del modelo y procesar casos con datos faltantes. Es a veces más fácil de interpretar que los modelos lineales cuando hay variables continuas y categóricas y tiende a involucrar menos variables que los modelos tradicionales.

La información se presenta mediante un diagrama en forma de árbol invertido, que se inicia con un nodo raíz que contiene todas las observaciones, este es dividido en dos subgrupos determinados por la partición de una variable elegida, generando dos nodos descendentes. Los dos subgrupos son divididos usando la dicotomización de una segunda variable, algunos de los nodos resultantes son terminales y asignados a la clase de mayor representación, mientras que los otros nodos continúan dividiéndose hasta llegar a un nodo terminal. CART considera todas las posibles particiones para todas las variables y para cada valor del rango observado. Cada partición es jerarquizada según un criterio de la "calidad de la partición".

El proceso de partición de los nodos culmina con la obtención del árbol máximo, CART procede a "podar" reduciendo los nodos terminales hasta llegar al árbol óptimo. La evaluación de cada subárbol se realiza mediante estimaciones de las tasas de mala clasificación (por sustitución, por muestras de validación o por validación cruzada). En la estimación por sustitución intervienen la probabilidad a priori y el costo de penalización definidos por el investigador.

CART entrega también un listado de particiones sustitutas y competidoras, las primeras tienen la cualidad de imitar la acción de las particiones primarias, permitiendo el trabajo con datos faltantes y las segundas presentan la mejor partición posible que reduce la impureza del nodo.

3. METODOLOGIA

La muestra utilizada en el análisis fue de 3150 estudiantes evaluados por EMTAJOVEN, con una media de edad de 13.7 ± 1.1 años y 50.6 % de mujeres.

Esta encuesta empleó un muestreo bi-etápico. En la primera etapa se seleccionaron los colegios que tienen de 6^o E.B. a 1^{ero} E.M. y en la segunda se seleccionaron aleatoriamente cursos de los colegios participantes. Todos los estudiantes de los cursos seleccionados fueron encuestados.

El factor de ponderación asociado a cada encuesta fue dado por: $W = w1*w2*f1*f2*f3*f4$

donde $w1$ = probabilidad inversa de seleccionar un colegio

$w2$ = probabilidad inversa de seleccionar una clase dentro del colegio

$f1$ = factor de ajuste de las no respuestas a nivel colegio calculado por categoría (pequeña, media, grande)

$f2$ = factor de ajuste al curso calculada por escuela

$f3$ = factor de ajuste de las no respuestas a nivel estudiante calculado por clase

$f4$ = factor de ajuste post-estratificación calculada por género y grado.

Además de este factor de ponderación, para el análisis se consideraron 26 variables, todas categóricas, relacionadas con características personales del estudiante, conocimiento y actitudes hacia el tabaco y para dejar de fumar, acceso y disponibilidad, exposición al humo de tabaco, familiaridad con propaganda en los últimos 30 días y curriculum escolar sobre el consumo y no consumo de tabaco en el último mes.

En el análisis estadístico se utilizaron los programas STATA 6.0 y la versión 4.0 de CART Salford Systems (1999, 2000), que incorpora la opción para usar las ponderaciones en la construcción de los árboles de clasificación.

Construcción del árbol de clasificación

Se especificaron las 25 variables categóricas, como independientes (Ver Tabla 1), más la variable ponderación. La variable haber fumado cigarrillo o consumido algún producto tabaco en el último mes (Fumador, No fumador) fue especificada como dependiente.

La validación del árbol se realizó a través de una muestra de prueba correspondiente al 40 % de los datos.

Se utilizó iguales probabilidades a priori y los costos de mala clasificación elegidos fueron las combinaciones de los valores 1, 1.5, 2, 2.5 y 3. En cada árbol obtenido se documentaron las tasas de especificidad, sensibilidad y clasificación correcta de las muestras de construcción y validación al nivel del 5 %. Del conjunto de árboles distintos, cuya prueba fue no significativa en las tres características, se eligió el que presentó la mayor especificidad a expensas de perder sensibilidad (criterio clínico) Gil (2000). Del árbol elegido se entregarán las particiones sustitutas que reemplazan a las particiones primarias en el caso de observaciones faltantes y la importancia relativa de las variables en el estudio, dada por la capacidad de una variable a ser sustituta de la partición primaria.

4. RESULTADOS

La prevalencia del consumo de tabaco en el último mes fue 45.1%. En la Tabla 1 se presenta la distribución porcentual de los estudiantes fumadores y no fumadores para las variables según característica y categorías consideradas en el estudio.

Tabla 1. Distribución porcentual (%) de los estudiantes fumadores y no fumadores evaluados por EMTAJOVEN 2000 según variables (rótulo) y categorías consideradas en el estudio.

VARIABLES	FUMADORES %	NO FUMADORES %
Características personales:		
1. Sexo (CR56):		
Hombres	44.1	53.7
Mujeres	55.9	46.3
2. Edad (CR552):		
≤ 12 años	14.9	18.9
13 años	23.8	32.4
14 años	29.2	29.8
15 años	23.1	15.1
16 y más	9.0	3.8
3. Curso (CLR57):		
6° E.B.	5.2	2.7
7° E.B.	31.7	39.3
8° E.B.	29.0	31.6
1° E.M.	26.4	26.4
4. Edad en que empezó a fumar (CRR2):		
2 = Menos de 10 años	17.1	7.0
3 = 10 -13 años	68.1	36.3
4 = Más de 14 años	10.5	5.8
Conocimientos y actitudes:		
5. No le importaría si el humo de su cigarrillo molesta a los demás (CLR60)		
1 = Verdadero	30.1	22.0
2 = Falso	69.2	78.1
6. El cigarrillo light es menos malo que el regular (CLR67)		
1 = Verdadero	37.8	16.8
2 = Falso	27.4	22.9
3 = No sé	34.8	60.3
7. Alguien de su familia le ha conversado sobre los efectos dañinos de fumar cigarrillo (CR17)		
1 = Si	78.7	84.4
2 = No	21.3	15.6
8. Cree que fumará cigarrillo en el próximo año (CRR18).		
1 = Intención negativa	27.3	82.1
2 = Intención positiva	72.7	17.9

9.	Cree que fumará cigarrillo en los próximos 5 años (CRR19)		
	1 = Intención negativa	43.5	81.3
	2 = Intención positiva	56.5	28.7
10.	Cree que los lolos que fuman tienen más o menos amigos (CR21).		
	1 = Más amigos	15.3	19.1
	2 = Menos amigos	5.8	10.6
	3 = No hay diferencia	78.9	70.3
11.	Cree que las lolas que fuman tienen más o menos amigas (CR22).		
	1 = Más amigos	15.5	20.3
	2 = Menos amigos	8.0	12.3
	3 = No hay diferencia	76.5	67.4
12.	El fumar te hace sentir más o menos "bacán" (CR23)		
	1 = Más bacán	47.4	30.6
	2 = Menos bacán	9.8	7.8
	3 = No hay diferencia	42.7	61.6
13.	Cree que fumar hace que los lolos se vean más atractivos (CR24).		
	1 = Más atractivos	15.6	11.1
	2 = Menos atractivos	15.5	32.2
	3 = No hay diferencia	68.9	56.7
14.	Cree que fumar hace que las lolas se vean más atractivas (CR25)		
	1 = Más atractivas	12.4	8.7
	2 = Menos atractivas	23.2	43.8
	3 = No hay diferencia	64.4	47.8
15.	Te han brindado ayuda o consejo para dejar de fumar? (CRR41)		
	1 = No he fumado	8.7	68.3
	1 = Sí, de un programa, amigo(a), familia o de todos ellos	64.9	20.0
	3 = No he recibido ayuda	26.4	11.7

Acceso y disponibilidad

16.	En que lugar fumas usualmente? (CRR12)		
	1 = No he fumado	8.3	74.7
	2 = En mi casa o colegio	23.3	5.3
	3 = En otro lugar distinto a la casa o colegio (casa amigo(a), lugares públicos, otro)	68.4	20.0
17.	Es posible comprar cigarrillos sueltos donde vive?(CLR59)		
	1 = Sí	2.8	78.3
	2 = No	9.9	8.5
	3 = No sé	7.3	13.2

Exposición al humo de tabaco

18.	En los últimos 7 días Alguien ha fumado en tu hogar u otros lugares en presencia tuya? (X1)		
	0 = ningún día	9.1	24.2
	1 = al menos un día	90.9	75.8
19.	Tus padres fuman? (CRR15)		
	1 = Ninguno de ellos	24.1	35.0
	2 = Al menos uno de ellos	74.7	63.6
	3 = No sé	1.2	1.5
20.	Casi todos sus compañeros de clase fuman? (CR61)		
	1 = Verdadero	42.5	31.0
	2 = Falso	15.4	20.0
	3 = No sé	42.2	49.1

21. Algunos de sus amigos(as) cercanos fuman cigarrillos? (CR28)		
1 = Ninguno(a)	5.2	18.5
2 = Algunos de ellos	37.7	60.0
3 = La mayoría o todos ellos	51.1	21.6

Familiaridad con propaganda relacionada a fumar cigarrillo en los últimos 30 días

22. Cuánta antipropaganda ha visto u oído? (CR42)		
1 = Mucha	30.3	24.9
2 = Poca	44.5	52.5
3 = Ninguna	25.3	22.6
23. Cuántas veces ha visto marcas de cigarrillos anunciada mientras veía deportes u otro programa de TV? (CR46)		
1 = Nunca veo TV	4.9	3.5
2 = Muchas veces	40.1	36.9
3 = A veces	46.4	49.7
4 = Ninguna vez	8.9	9.9
24. Cuántas veces ha visto marcas de cigarrillos en avisos en la calle o periódicos o revistas? (X2)		
0 = Ninguna	1.7	1.0
1 = Muchas veces	22.8	25.3
2 = Pocas veces	75.5	73.7

Currículum escolar

25. Durante el presente año ha tenido alguna clase sobre el peligro o sobre el por qué los jóvenes fuman o sobre los efectos de fumar cigarrillo. (X3)		
1 = Sí	18.5	21.5
2 = No	71.4	67.4
3 = No estoy seguro(a)	10.1	11.0

Construcción del árbol de clasificación con factor de ponderación

Sobre la base de la metodología, se construyeron 25 árboles de clasificación usando las combinaciones de los costos de mala clasificación propuestos (C_1 = costo de clasificar mal a un estudiante que fuma y C_2 = costo de clasificar mal a un estudiante que no fuma). Resultaron 13 árboles diferentes y en cada uno de ellos se docimaron las tasas de especificidad, sensibilidad y clasificación correcta de la muestra de construcción y validación expandida. Cinco árboles presentaron diferencias no significativas en las tres características al nivel de 5 %, de estos, se eligió el árbol de clasificación con 9 nodos terminales por presentar la tasa de especificidad más alta (Ver Tabla 2).

La secuencia de podaje del árbol de clasificación elegido, con igual probabilidad a priori para las categorías de la variable dependiente, e iguales costos de mala clasificación se presenta en la Tabla 3.

En la secuencia de podaje se muestra que el árbol máximo cuenta con 112 nodos terminales con un costo relativo sobre la muestra de validación de $42.6\% \pm 0.002$ desviaciones estandar, respecto del costo inicial de mala clasificación (50%). Además se observan los costos relativos de mala clasificación sobre la muestra de construcción en cada uno de los subárboles.

El Gráfico 1 muestra el árbol de clasificación con 9 nodos terminales, resultado de particiones de las siguientes variables: ¿En qué lugar fumas usualmente? (CR12, nodo raíz), ¿Crees que fumarás cigarrillo el próximo año? (CRR18), No me importaría si el humo de cigarrillo molestara a los demás? (CLR60), El cigarrillo light es menos malo que el regular (CLR67), ¿En qué curso estás? (CLR57), ¿Cuál es tu sexo? (CR56), ¿Crees que los lolos que fuman tienen más o menos amigos? (CR21), ¿Cuántos años tienes? (CR552). En cada uno de los nodos terminales se observa el total de estudiantes en la muestra expandida y la composición de acuerdo a la clase de pertenencia, el grupo con mayor representación determina la clase asignada al nodo.

Tabla 2. Comparación de las tasas de especificidad, sensibilidad y clasificación correcta de las muestras de validación y construcción y valor p-value para los 13 árboles obtenidos según costos de mala clasificación propuestos.

Arbol	Número nodos terminales	Costos	Tasas muestras:			p-value
				Construcción	Validación	
1	9**	C ₁ = 1.0 C ₂ = 1.0	Especificidad: Sensibilidad: Clasificación correcta:	0.800 0.893 0.841	0.795 0.871 0.830	0.733 0.059 0.415
2	20	C ₁ = 1.0 C ₂ = 1.5	Especificidad: Sensibilidad: Clasificación correcta:	0.859 0.865 0.862	0.831 0.823 0.828	0.033 0.001 0.009
3	43	C ₁ = 1.0 C ₂ = 2.0	Especificidad: Sensibilidad: Clasificación correcta:	0.934 0.803 0.862	0.872 0.730 0.806	0.000 0.000 0.000
4	13	C ₁ = 1.0 C ₂ = 2.5	Especificidad: Sensibilidad: Clasificación correcta:	0.936 0.679 0.823	0.905 0.658 0.790	0.001 0.221 0.021
5	12	C ₁ = 1.0 C ₂ = 3.0	Especificidad: Sensibilidad: Clasificación correcta:	0.942 0.655 0.815	0.912 0.642 0.786	0.001 0.455 0.045
6	4*	C ₁ = 1.5 C ₂ = 1.0	Especificidad: Sensibilidad: Clasificación correcta:	0.732 0.927 0.818	0.741 0.920 0.824	0.576 0.468 0.668
7	35	C ₁ = 1.5 C ₂ = 2.0	Especificidad: Sensibilidad: Clasificación correcta:	0.892 0.862 0.879	0.846 0.802 0.826	0.000 0.000 0.000
8	67	C ₁ = 1.5 C ₂ = 2.5	Especificidad: Sensibilidad: Clasificación correcta:	0.940 0.846 0.899	0.866 0.767 0.820	0.000 0.000 0.000
9	6*	C ₁ = 2.0 C ₂ = 1.0	Especificidad: Sensibilidad: Clasificación correcta:	0.729 0.931 0.818	0.736 0.923 0.823	0.665 0.621 0.721
10	2*	C ₁ = 2.0 C ₂ = 1.5	Especificidad: Sensibilidad: Clasificación correcta:	0.738 0.918 0.818	0.750 0.913 0.826	0.452 0.621 0.567
11	25	C ₁ = 2.0 C ₂ = 2.5	Especificidad: Sensibilidad: Clasificación correcta:	0.862 0.875 0.868	0.836 0.821 0.829	0.045 0.000 0.003
12	12*	C ₁ = 2.5 C ₂ = 1.0	Especificidad: Sensibilidad: Clasificación correcta:	0.712 0.942 0.814	0.717 0.933 0.818	0.762 0.305 0.777
13	17	C ₁ = 3.0 C ₂ = 1.0	Especificidad: Sensibilidad: Clasificación correcta:	0.701 0.961 0.815	0.695 0.937 0.808	0.720 0.002 0.623

*Árboles de clasificación con tasas no significativas en las tres características.

**Árbol de clasificación elegido.

Tabla 3. Secuencia de podaje del árbol de clasificación elegido. Costos relativos de mala clasificación en la muestra de prueba o validación y costos relativos de sustitución.

Número de nodos terminales	Costos relativos en la muestra de validación	Costos relativos de sustitución
112	0.426 ± 0.002	0.183
25	0.346 ± 0.002	0.260
22	0.343 ± 0.002	0.267
18	0.337 ± 0.002	0.276
17	0.344 ± 0.002	0.279
15	0.352 ± 0.002	0.284
14	0.348 ± 0.002	0.287
11	0.335 ± 0.002	0.299
9*	0.335 ± 0.002	0.307
2	0.337 ± 0.002	0.344
1	1.000 ± 0.000	1.000

Tabla 4. Evaluación del árbol de clasificación en la muestra expandida de validación y construcción.

Muestra validación: Especificidad = 0.795 Sensibilidad = 0.871 Clasificación Correcta = 0.830			
Clasificación actual	Clasificación predicha		Total
No Fumador	No fumador 46966.57	Fumador 12131.89	59098.46
Fumador	6683.19	45009.96	51693.15
Muestra validación: Especificidad = 0.800 Sensibilidad = 0.893 Clasificación Correcta = 0.841			
Clasificación actual	Clasificación predicha		Total
No fumador	No fumador 77292.95	Fumador 19287.91	96580.85
Fumador	8177.62	679441.88	76119.49

En la Tabla 4 se presenta la evaluación del árbol de clasificación final, en la muestra expandida de validación y construcción respectivamente.

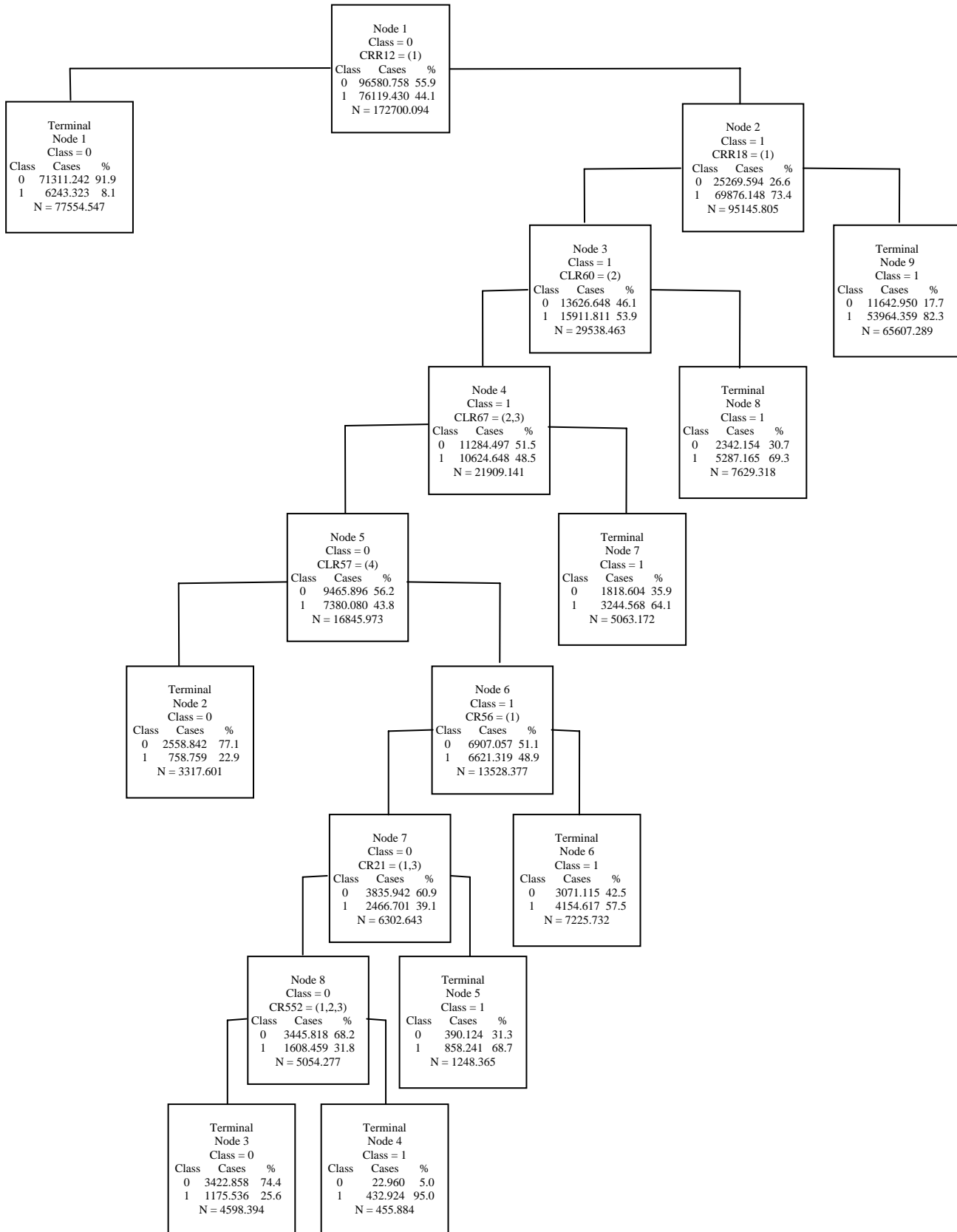
La Tabla 5 muestra las particiones sustitutas, las que pueden ser usadas en caso de valores faltantes en las variables de las particiones primarias y el valor de asociación que indica la capacidad de la variable en imitar la partición original.

Tabla 5. Particiones de las variables sustitutas y valor de asociación.

Nodo	Variable	Partición*	Asociación
1	CRR41	1 s	0.775
	CRR2	1 s	0.578
	CRR18	1 s	0.462
	CRR19	1 s	0.205
	X1	0 s	0.145
2	CR24	2 s	0.047
	CRR19	1 s	0.023
	CR21	2 s	0.021
	CR22	2 s	0.017
5	CR552	4,5 s	0.281
6	CR23	1 s	0.105
	CR25	2 s	0.098
	CR24	2 s	0.087
	CR22	1,2 s	0.068
	CR21	2 s	0.062
7	CR22	1,3 s	0.520
	X2	1,2 s	0.079
	CR23	1,3 s	0.032

*partición estandar = s

Gráfico 1. Arbol de clasificación final con factor de ponderación basado en N = 3014 estudiantes de la región metropolitana evaluados con EMTAJOVEN 2000. Cada nodo terminal está asignado a la clase fumadores (1) o no fumadores (2)



En la Tabla 6 se presenta la importancia relativa de las variables. Cada variable es ranqueada sobre la base de la reducción de la impureza sobre los nodos y su participación como sustituta. La variable CRR12, lugar dónde fuma usualmente, es la que presenta el mayor valor, se le asigna el 100 %, le sigue en importancia la variable ¿Te han brindado ayuda o consejo para dejar de fumar? (CRR41).

Tabla 6. Importancia relativa de las variables consideradas en el estudio.

Variables	Importancia relativa %
CRR12	100.0
CRR41	80.51
CRR18	72.27
CRR2	61.24
CRR19	39.38
X1	10.15
CR24	2.30
CLR60	1.39
CR552	1.18
CR21	0.97
CLR57	0.97
CR56	0.94
CLR67	0.88
CR22	0.61
CR25	0.35
CR23	0.18
CR42	0.00

5. CONCLUSIONES

El árbol de clasificación con factor de ponderación determinó un modelo para describir distintos grupos de estudiantes que han fumado alguna vez cigarrillo o han usado algún producto tabaco en el último mes, con una tasa de especificidad de 80 %, 89 % de sensibilidad y 84 % de clasificación correcta sobre la muestra de construcción expandida. Los grupos determinados por el árbol de clasificación están relacionados al lugar dónde fuman usualmente (CRR12); relacionados al conocimiento y actitud de los estudiantes, tales como ¿Crees que fumarás el próximo año? (CRR18), No le importa si el humo del cigarrillo molesta a los demás (CR60), El cigarrillo light es menos malo que el corriente (CLR67), ¿Crees que los lolos que fuman tienen más o menos amigos? (CR21); y relacionada a características personales, como curso o grado del estudiante (CLR57), sexo (CR56) y edad (CR552).

Las variables que pueden ser utilizadas en caso de valores faltantes en las variables primarias, están relacionadas a dejar de fumar (CRR41), relacionada a la exposición al humo de tabaco que otros fuman (X1), relacionada a la actitud hacia el tabaco en el próximo año y dentro de 5 años (CRR18 y CRR19) y la edad en que empezó a fumar (CRR2) y la edad cronológica (CR552), entre otras variables.

Estos resultados podrían ayudar en la búsqueda de determinantes en la implementación de un programa de prevención y control del tabaco en la Región Metropolitana.

REFERENCIAS

- BREIMAN, L.; J. FREIDMAN; R. OLSHEN and CH. STONE (1997): CART, A Salford Systems Implementation of the Original Program Interface and Documentation (D. Steinberg and P. Colla). Salford Systems Inc.
- GIL LAVERDE, J. F. (2000): Desarrollo y validación de un modelo de predicción del estado vital de pacientes admitidos a unidades de cuidado intensivo en Colombia. Tesis para optar al grado de Mag. en Bioestadística, Chile.

PAZ CIUDADANA (1999): "Consumo de alcohol, tabaco y drogas en la población escolar de Chile. MINSAL, MINEDUC", Consejo Nacional de control de estupefaciente, **Rev. Pediatría al día** 15, 246-247

EMTA JOVEN (2000): Proponen estrategias para evitar tabaquismo juvenil, Diario "La Tercera", 31 de mayo.

SALFORD SYSTEMS (1999): Data mining with Decision Trees: An Introduction to CART. Salford Systems Inc.

SALFORD SYSTEMS (2000): CART for Windows. User's Guide. Salford Systems Inc.