



FACULTAD  
DE CIENCIAS  
ECONÓMICAS



Universidad  
Nacional  
de Córdoba

# REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

## Uso de árboles de clasificación para la predicción de empresas con dificultades financieras en mercados latinoamericanos

Norma Patricia Caro, Mariana Guardiola, Pablo Ortiz

Capítulo del Libro Métodos estadísticos aplicados para diagnosticar y predecir el estado de crisis financiera en empresas que cotizan en mercados latinoamericanos, 1º ed.

publicado en 2017 - ISBN 978-987-3840-55-5



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

## Capítulo 4

# Uso de árboles de clasificación para la predicción de empresas con dificultades financieras en mercados latinoamericanos<sup>6</sup>

**Norma Patricia Caro**

**Mariana Guardiola**

**Pablo Ortiz**

### 4.1. Introducción

La predicción de la insolvencia es uno de los temas centrales del análisis financiero, y ha suscitado el interés no sólo del ámbito académico, sino también de un amplio abanico de usuarios relacionados con el mundo empresarial. A partir de la década de 1960 (Beaver, 1966, 1968; Altman, 1968) se han aplicado diferentes metodologías, tanto univariadas como multivariadas (discriminante, logit, probit, etc.), en las que los ratios contables funcionan como variables explicativas, obteniendo resultados muy satisfactorios.

Altman (1993) utilizó árboles de clasificación (decisión) para la predicción de crisis empresarial en economías desarrolladas. En esta línea, el presente capítulo extiende este estudio a empresas que cotizan sus acciones en las bolsas de mercados latinoamericanos, mediante un modelo de predicción de insolvencia a fin de detectar qué ratios contables influyen primordialmente en las empresas en crisis. De esta manera se estudia la relación entre el estado de crisis de una empresa y el comportamiento de ciertos ratios contables. A partir de este estudio, se propicia que empresas, entidades financieras, inversores, entre otros, puedan predecir el estado de vulnerabilidad financiera de las empresas según el comportamiento de algunos de los indicadores contables, al

---

<sup>6</sup> Adaptación del artículo “Árboles de clasificación como herramienta para diagnosticar dificultades financieras a través de los ratios contables en empresas latinoamericanas”. Autores: Caro, N; Guardiola, M y Ortiz, P. En prensa. Revista de Estudios Económicos y Empresariales – México. Aceptado en 2017.

tiempo que constituye una motivación a continuar investigando en esta área de interés.

La determinación de puntos de corte a través de la aplicación de la técnica de árboles de decisión puede ser considerada como paso intermedio para la aplicación de modelos para datos longitudinales, cuyos primeros antecedentes en relación a su aplicación en países latinoamericanos, se encuentran en Caro, et al. (2013) y Caro y Díaz (2015). De esta manera, estos puntos de corte permiten analizar los estados contables de las empresas y tomar decisiones en función del valor que asumen los mismos en cada ejercicio contable anual.

Así, la presente investigación contribuye a construir modelos de predicción que permitan una primera aproximación para dilucidar qué ratios financieros resultan significativos en la clasificación de las empresas, al tiempo que establece una regla que determina los umbrales a partir de los cuales una nueva empresa está en una situación de crisis o no.

## **4.2. Muestra y variables**

Para el análisis de las empresas que cotizan en los distintos mercados latinoamericanos se utilizó la información de los estados contables disponibles en las respectivas Bolsas en la década del 2000. En el caso de Argentina, la muestra está conformada con la información de los estados contables anuales de 57 empresas, 44 sanas y 13 en crisis. En Chile, se analizaron 191 empresas, 158 sanas y 33 en crisis. Por último se utilizó información de 50 empresas de Perú, 37 de ellas sanas y 13 en crisis. Para las empresas en crisis se tomó el estado contable anual anterior al año en que se detectó su vulnerabilidad financiera y para las sanas, el del último período disponible. Los motivos para que sean consideradas en crisis se expusieron en el capítulo 2 del presente libro.

En los tres mercados estudiados, la variable dependiente es el estado de la empresa, con dificultades financieras (crisis) y sin ellas (sanas). Por su parte, las variables predictoras se encuentran definidas por ratios contables, los cuales son indicadores construidos a partir de la información de los estados contables presentados por las empresas, que han sido empleados en estudios previos.

Los indicadores considerados en este capítulo son los que utiliza, Caro, Díaz, et al. (2013) basados en Altman (1968) y Jones y Hensher (2004), siendo los ratios más utilizados y que han sido validados en múltiples aplicaciones en diferentes mercados:

- **Ratio de Flujo de Fondos (FF\_AT):** mide la posición de efectivo originado por las actividades operativas respecto al total del activo.
- **Ratio de Liquidez (E\_AT):** mide la proporción de los recursos más líquidos de la empresa (disponibilidades e inversiones que pueden hacerse efectivas dentro de las 24 horas) sobre el total de activo.
- **Ratio de Rotación del Activo Total (V\_AT):** es el coeficiente que mide la relación entre las ventas y el activo total.
- **Ratio de Endeudamiento (D\_PN):** mide la relación de las deudas (capital de terceros) sobre el capital aportado por los dueños de la empresa.
- **Ratio de Capital de trabajo (CT\_AT):** expresa el grado de fluidez de los activos e indica en qué medida la inmovilización del activo total resulta neutralizada por el capital de trabajo.
- **Ratio de Rentabilidad económica (GE\_AT):** mide la rentabilidad en función de las ganancias de la explotación.
- **Tamaño de la empresa (LNA):** medido como el logaritmo natural del activo total.

### 4.3. Metodología

Altman (1993) fue uno de los iniciadores en el uso de metodologías clásicas (como el análisis discriminante) y no clásicas (como las redes neuronales y los árboles de clasificación y/o decisión) en economías desarrolladas.

En este capítulo se aplica una metodología basada en el algoritmo de inducción de reglas denominado CHAID (*Chi-squared Automatic Interaction Detector*), que fue desarrollado por Kass (1980) y Breiman et al. (1984). Esta técnica estadística de segmentación genera un árbol de reglas que describe los distintos segmentos en que se divide la muestra en relación a la variable dependiente, permitiendo clasificar a las empresas de acuerdo a los diferentes valores que presentan sus ratios contables.

Aunque son muchos los algoritmos capaces de generar reglas basados en árboles de clasificación, la elección del CHAID se basa tanto por su amplia difusión como por sus importantes beneficios.

En primer lugar, no está basado en ninguna distribución de probabilidad, empleando solamente los test de bondad de ajuste Chi-cuadrado

(derivados de tablas de contingencia) que, con un tamaño muestral aceptable, casi siempre produce buenos resultados. En segundo lugar, también permite determinar una variable a maximizar, lo que es deseable y no siempre posible con otras técnicas de segmentación. En tercer lugar, la clasificación mediante segmentos es relativamente fácil de analizar, ya que aporta reglas intuitivas sencillas de interpretar por usuarios no expertos, algo que no siempre ocurre cuando se trabaja con otras metodologías. En cuarto lugar, la técnica asegura que los segmentos siempre tienen significado estadístico, es decir, que son todos diferentes y los mejores posibles en base a los datos muestrales. Consecuentemente, las clasificaciones realizadas con las reglas encontradas son mutuamente excluyentes, y por tanto el árbol de decisión proporciona una única respuesta basada en el cálculo de las probabilidades de pertenecer a una cierta clase. Finalmente, CHAID es un algoritmo no binario, es decir, capaz de construir más de dos divisiones de los datos de acuerdo a las categorías a explicar en cada nodo.

El algoritmo CHAID consta de varios pasos en su aplicación hasta la obtención de las reglas (Breiman et al., 1984):

1. División de las variables explicativas continuas: Estas variables son discretizadas y divididas en un conjunto de categorías.
2. Fusión de categorías de las variables explicativas: Este algoritmo fusiona aquellos valores de las variables explicativas que son estadísticamente homogéneos respecto a la variable dependiente, y mantiene separados aquellos otros que son heterogéneos. Cada categoría final obtenida de una variable explicativa  $X$  representa un nodo hijo, si luego la variable es utilizada para dividir el árbol. Para cada variable explicativa  $X$ , el algoritmo encuentra la pareja de categorías de  $X$  que tienen menos diferencias significativas con respecto a la variable dependiente  $Y$ , es decir, con mayor  $p$ -value, el cual se obtiene a partir del test Chi-cuadrado.
3. División de nodos: cada variable es evaluada en su asociación con la variable dependiente, según el  $p$ -value obtenido con el test estadístico, de modo que el algoritmo selecciona el mejor predictor para formar la primera división del árbol, es decir, la variable explicativa con mayor asociación con la variable dependiente (aquella cuyo test Chi-cuadrado proporcionó el menor  $p$ -value). Si este valor es menor o igual que el umbral de división fijado por el usuario, entonces la variable se utiliza como variable de división para el nodo en cuestión, y cada una de las categorías fusionadas de la variable de división define un nodo hijo. Una vez dividido el nodo en cuestión, los nodos hijos son examinados para ver si permiten más divisiones

con la aplicación de este proceso de fusión/división. Este proceso continua recursivamente hasta que el árbol queda configurado y no se pueden realizar más divisiones. El nivel de confianza de cada regla (nodo terminal) representa la proporción de registros de cada regla que pertenecen a la categoría  $j$  seleccionada, mientras que de manera similar, el nivel de confianza de un conjunto de reglas puede definirse como la proporción de casos de cada regla que pertenecen a la categoría  $j$  dada.

4. Índice: el índice de cada una de las reglas obtenidas para una categoría  $j$  se obtiene como el ratio entre el nivel de confianza de cada regla o nodo terminal y el nivel de confianza de la categoría  $j$  en la muestra total. La ganancia para cada nodo terminal ( $t$ ) se define en términos absolutos como el número de casos en la categoría  $j$  seleccionada. Para un conjunto de reglas o nodos terminales, y en términos relativos, la ganancia representa el porcentaje de casos en la categoría  $j$ , y si es igual a 1, el caso posee la categoría  $j$ , y 0 en caso contrario.

#### 4.4. Resultados<sup>7</sup>

En los tres mercados la rentabilidad resultó determinante con una fuerte incidencia en la clasificación de las empresas. Posteriormente, a efectos de analizar el impacto de otros ratios que se veían soslayados, se la excluyó, obteniendo resultados relativamente más disimiles entre los países.

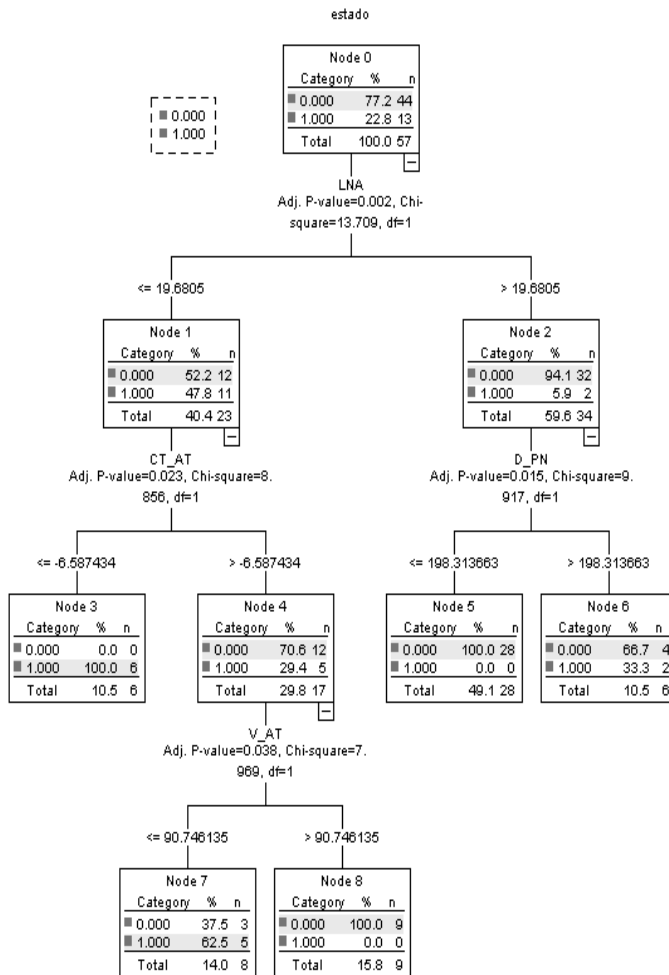
En el caso de Argentina, donde se obtuvo el árbol de mayor complejidad en cuanto a la cantidad de ramificaciones (Figura 4.1), el tamaño fue la principal variable predictora para el estado de las empresas. El 94,1% de empresas de mayor tamaño (LNA), con un ratio superior a 19.68%, son sanas. Adicionalmente, de este grupo de empresas, el 100% con un ratio de endeudamiento (D\_PN) inferior a 198.31% son sanas.

Por otra parte, las empresas de menor tamaño se encuentran distribuidas de manera bastante equitativa entre las categorías, siendo 47.8% en crisis; aunque de éstas, el 100% que tienen un ratio de capital de trabajo (CT\_AT) inferior a -6.59% se corresponden a esa categoría. En cuanto a la totalidad de empresas que devienen con estas características tienen una rotación de activos (V\_AT) menor a 90.75%.

---

<sup>7</sup> Procesados con SPSS Statistics.

**Figura 4.1. Árbol de Clasificación - Argentina**

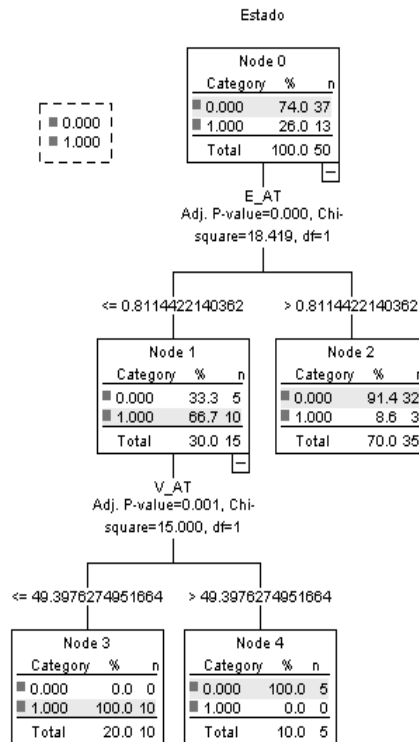


Fuente: elaboración propia

En cuanto a Perú, la liquidez resultó la variable predictor más relevante (Figura 4.2), el 91.4% de empresas con un ratio superior a 0.81% son sanas, mientras que el 66.7% con un ratio inferior a ese valor son empresas en crisis. Adicionalmente, este grupo de empresa de menor liquidez queda totalmente determinado por la rotación de activos, puesto

que, la totalidad de empresas con este ratio inferior a 49.4% están crisis, mientras que las que superan esa magnitud son sanas.

**Figura 4.2.** Árbol de Clasificación - Perú

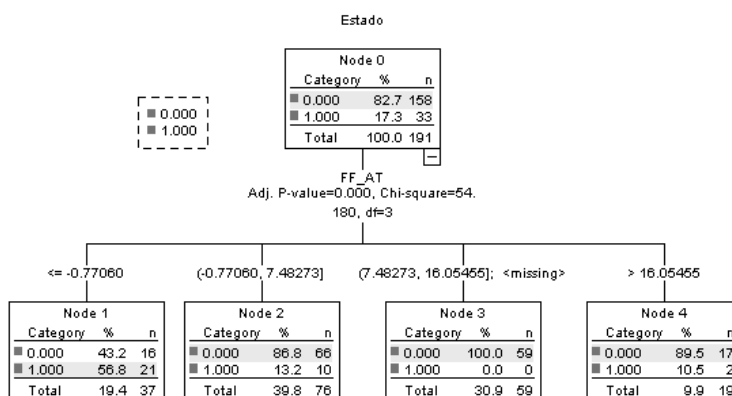


Fuente: elaboración propia

Con respecto a Chile, el flujo de fondos resultó la variable predictora dominante (Figura 4.3). De la totalidad de empresas en crisis de la muestra de este país, el 64% tienen un nivel de flujo de fondos (FF\_AT) inferior a -0.77% (constituyendo el 56.8% de las empresas con esta característica), mientras que el 92% de empresas con un ratio superior a esa magnitud, son empresas sanas.



**Figura 4.3.** Árbol de Clasificación - Chile



Fuente: elaboración propia

En relación a la bondad del funcionamiento del modelo en los diferentes países, considerando el porcentaje de clasificación correcta (Tabla 4.1), Perú fue el de mejor performance tanto a nivel general (94%) como en la clasificación específica de las empresas sanas (100%), alcanzando además 76.9% en el caso de empresas en crisis. Seguidamente, las empresas de Argentina fueron clasificadas correctamente en un 91.2%, junto al mejor desempeño en cuanto a las empresas en crisis, con un 84.6% de clasificación correcta. Finalmente, la tasa de clasificación global de las empresas chilenas fue muy buena, alcanzando un 85.3%, aunque en relación a las empresas en crisis la performance resultó un poco inferior, 63.6%.

**Tabla 4.1.** Porcentajes de clasificación correcta por mercado

Estado	Argentina	Chile	Perú
Empresas sanas (0)	93,2%	89,9%	100%
Empresas en crisis (1)	84,6%	63,6%	76,9%
<b>Porcentaje Global</b>	<b>91,2%</b>	<b>85,3%</b>	<b>94%</b>

Fuente: elaboración propia

Las reglas obtenidas identifican perfiles de alta o baja performance, lo que indicaría que si una determinada empresa posee cierto comportamiento en sus ratios, podría estar en peligro de crisis financiera. En función de los árboles de clasificación obtenidos (Figuras 4.1 a 4.3) se calcularon los intervalos de los diferentes ratios a partir de los cuales empresarios, bancos o inversores, pueden determinar en qué situación financiera se encuentra la empresa de interés (Tabla 4.2).

**Tabla 4.2.** Intervalos para predecir un estado de crisis financiera por mercado

<b>Ratios</b>	<b>Chile</b>	<b>Perú</b>	<b>Argentina</b>	
Tamaño	---	---	≤ 19,68	---
Capital de trabajo	---	---	≤ -6,59	> -6,59
Liquidez	---	≤ 0,81	---	---
Rotación de Activos	---	≤ 49,4	---	≤ 90,74
Flujo de Fondos	≤ -0,77	---	---	---

Fuente: elaboración propia

De la lectura de la tabla precedente se desprende que una empresa chilena tiene una alta posibilidad de presentar problemas financieros si su ratio referido al flujo de fondos asume un valor menor o igual a -0,77. Mientras que en el caso de las empresas peruanas, la probabilidad de atravesar procesos de crisis es mayor si su ratio de liquidez no supera el valor 0,81 y la rotación del activo es a lo sumo 49,4. Respecto a las empresas argentinas con mayores probabilidades de experimentar procesos de vulnerabilidad financiera, se pueden identificar dos tipos: empresas cuyos ratios de tamaño y capital de trabajo arrojen valores menores o iguales a 19,68 y -6,59, respectivamente y empresas con un índice de capital de trabajo superior a dicha magnitud pero con rotación del activo no mayor a 90,74.

## **4.5. Conclusiones**

En esta investigación se han presentado las principales características económico-financieras de las empresas cotizantes en las Bolsas latinoamericanas que han formado parte de la misma. Las reglas o perfiles obtenidos nos han permitido caracterizar a las empresas de acuerdo a los ratios de sus estados contables. Sin embargo, el principal valor agregado de este estudio no ha sido confirmar empíricamente las variables principales a tener en cuenta en la planificación financiera, sino determinar los niveles o intervalos cuantitativos a marcar como objetivo en cada una de ellas, propiciando la previsión de inconvenientes financieros.

Se concluye así que la aplicación de estos métodos estadísticos permite identificar empresas con problemas financieros, lo cual cobra relevancia en la modelación y predicción de este tipo de riesgo.

De este modo, en el presente capítulo se ha podido recomendar una regla que permite, con adecuados valores en los ratios, predecir un estado de vulnerabilidad para que cada empresa, atendiendo a sus circunstancias particulares (internas y externas), pueda realizar su planificación financiera de acuerdo a las reglas que considere alcanzables y sostenibles.