

**UNIVERSIDAD NACIONAL AGRARIA LA MOLINA**  
**FACULTAD DE ECONOMÍA Y PLANIFICACIÓN**  
**Departamento Académico de Estadística e Informática**



Trabajo monográfico

PERFIL DE LOS CLIENTES QUE ACEPTAN UNA TARJETA DE  
CRÉDITO DE UN BANCO VIA CALL CENTER UTILIZANDO EL  
ALGORITMO CHAID EXHAUSTIVO

Presentado para optar el título de Ingeniero Estadístico e Informático

DIANA ROSA ACOSTA PIZARRO

Modalidad de Examen Profesional

LIMA-PERÚ  
2014

A mispadres, a mi esposo Joséy a  
nuestra pequeña Luanapor apoyarme y  
motivarme en cada momento de mi vida y  
a todos aquellos que me apoyaron directa o  
indirectamente en la elaboración de este proyecto.

# ÍNDICE

RESUMEN	
I. INTRODUCCIÓN	1
II. PROBLEMA DE INVESTIGACIÓN	3
2.1. Fundamentación del problema de investigación	3
2.2. Formulación de las preguntas de investigación	4
2.3. Objetivos de la investigación	4
2.4. Justificación de la investigación	4
III. MARCO TEÓRICO	6
3.1. Árboles de Clasificación	6
3.1.1. Definición de Árboles de Clasificación	6
3.1.2. Elementos de los árboles de clasificación	7
3.1.3. Fases de los árboles de clasificación	7
3.1.5. Ventajas y desventajas de los Árboles de clasificación	9
3.1.6 Descripción de algoritmos de Árboles de Clasificación	10
3.2. Método CHAID Exhaustivo	11
3.2.1 Definición del algoritmo CHAID Exhaustivo	11
3.2.2 Creación del algoritmo CHAID Exhaustivo	11
3.2.3 Precisión del algoritmo CHAID Exhaustivo	13
3.2.4 Ventajas y desventajas del método	13
IV. METODOLOGÍA DE LA INVESTIGACIÓN	15
4.1. Tipo de investigación	15
4.2. Formulación de la Hipótesis	15
4.3. Identificación de variables	15
4.4. Población y muestra	19
4.5. Instrumento de colecta de datos	19
V. PROCEDIMIENTO DE ANÁLISIS DE DATOS	20
VI. RESULTADOS	22
6.1. Análisis Univariado y Bivariado	22
6.2. Árbol de Clasificación CHAID Exhaustivo	24
6.3. Comparación entre el modelo elegido y resultados reales	28
VII. CONCLUSIONES	30
VIII. RECOMENDACIONES	31
IX. BIBLIOGRAFÍA	32
X. ANEXO	33

## **Índice de Cuadros y Gráficos**

Gráfico N° 1: Fases de aprendizaje y clasificación	8
Cuadro N° 1: Descripción de las variables	18
Cuadro N° 2: Tabla de frecuencias de las variables cualitativas	22
Cuadro N° 3: Tabla de contingencias de las variables	23
Cuadro N° 4: Tabla de Resumen del modelo	24
Gráfico N° 2: Árbol de Clasificación-Método CHAID EXHAUSTIVO	25
Cuadro N° 5: Ganancias para los nodos: ACEPTA TC	26
Cuadro N° 6: Nodos con mayor índice de Ganancias	27
Cuadro N° 7: Tabla de riesgos del modelo	27
Cuadro N° 8: Tabla de Clasificación correcta del modelo	28
Cuadro N° 9: Escenario real de aceptación de tarjetas de crédito	29
Cuadro N° 10: Escenario de pronóstico de aceptación de tarjetas de crédito usando el modelo	29

## **RESUMEN**

El presente estudio tiene como objetivo principal identificar el perfil de los clientes del departamento de Lima que aceptan una tarjeta de crédito de una entidad financiera cuando el producto es ofrecido por el canal de ventas Call Center.

Se utilizó la técnica de Árboles de Clasificación CHAID Exhaustivo el cual proporciona buenos resultados de clasificación correcta de los clientes que aceptan una tarjeta de crédito vía Call Center.

Se consideró una muestra en un período de cinco meses (Diciembre 2013 a Abril 2014) logrando identificar que las variables más significativas que aportan en el modelo son la edad, el ingreso neto mensual y el tipo de tarjeta que se le ofrece al cliente. Estas variables presentan importancia relevante en el cliente para tomar la decisión de aceptar una tarjeta de crédito.

Los resultados obtenidos mediante el algoritmo CHAID Exhaustivo permitieron identificar los patrones que definen el perfil de los clientes que aceptan una tarjeta de crédito vía Call center con el fin de ser más efectivos, aumentando el número de ventas, reduciendo el número de llamadas, minimizando costos y tiempo.

## I. INTRODUCCIÓN

Para el sistema financiero cada vez es más importante conocer a sus clientes, saber a quienes dirigirse, y más aun a qué tipo de cliente ofrecer específicamente un producto acorde a su comportamiento y estilo de vida. El conocimiento del cliente ha sido desde los años 70's la forma en la que entidades financieras han podido gestionar y hacer estrategias de captación de los mismos.

Las entidades financieras están en una situación privilegiada para conocer las transacciones de sus clientes, así como cierta información relevante sobre sus características socioeconómicas y demográficas lo que les permite establecer perfiles de sus clientes.

Entre los principales productos financieros activos que se ofrecen se destaca la tarjeta de crédito. Mediante un sistema de reglas propuestas por expertos le asignan a cada cliente un cupo de endeudamiento. El sistema informático de la entidad está programado para no dejar que la persona gaste más de lo que su perfil le permite. Para poder obtener mayores colocaciones de tarjetas de crédito, las entidades financieras se centran no solo en conocer el perfil transaccional del cliente ya una vez que tienen su tarjeta de crédito sino que les es importante conocer el perfil del potencial cliente que aceptaría su producto. Es decir a quienes dirigirse, que tipo de tarjeta, línea de crédito y a través de que canal de venta ofrecérselo.

Las tecnologías de la información y la comunicación han permitido que parte del trabajo o procesos de las empresas se pueda desarrollar en diferentes lugares, y más aún, que sea hecho por otras empresas especializadas, generalmente a un costo menor o con mayor eficiencia o mayor eficacia. Este es el caso de una empresa de Outsourcing quien brinda a sus clientes servicios de tercerización en procesos integrales (BPO, Business Process Outsourcing). A esta empresa se le encarga el proceso de venta de tarjetas de crédito para una entidad financiera mediante llamadas telefónicas, por lo cual, instala un Call Center con los recursos óptimos y suficientes para la correcta gestión de la cartera de clientes potenciales.

La entidad financiera tiene la responsabilidad de enviar la base de datos donde mensualmente se trabaja la cartera de clientes, el supervisor es el responsable de realizar las estrategias de segmentación de la base que mejor le sean convenientes y las maneja a su criterio para poder llegar a la meta trazada por el cliente y que genere ganancias para la organización.

Por ello, para la empresa de Outsourcing que ofrece el servicio de Call Center le surge la necesidad de mejorar las estrategias de contactabilidad y aceptación de tarjetas de crédito para conocer a qué tipo de cliente se le está ofreciendo la tarjeta.

Por lo tanto, el objetivo del presente estudio se basa en identificar el perfil de los clientes del departamento de Lima que aceptan una tarjeta de crédito cuando se les ofrece por el canal de ventas Call Center. Para ello la técnica estadística de los Árboles de Clasificación utilizando el algoritmo CHAID Exhaustivo permitirá determinar si las variables utilizadas por la entidad financiera tienen capacidad discriminante en la aceptación de una tarjeta de crédito por parte de un cliente potencial.

El alcance de este estudio se limita a la evaluación de 40,471 clientes del departamento de Lima incluidos en las bases de datos enviadas por la entidad financiera en los últimos 5 meses (Diciembre 2013 a Abril 2014) que tienen una tarjeta de crédito pre-aprobada por la entidad financiera.

## **II. PROBLEMA DE INVESTIGACIÓN**

### **2.1. Fundamentación del problema de investigación**

Como se mencionó, el canal de ventas telefónicas Call Center es el responsable de contactarse directamente con cada cliente potencial a fin de poder ofrecerles una tarjeta de crédito. La cartera de clientes potenciales proviene del banco.

La entidad financiera envía una base de datos de clientes cada 15 días los cuales forman parte de la base total mensual que se toma en cuenta para los resultados mensuales de la gestión final y para los cuales el banco calcula las comisiones de pago por cada tarjeta activada.

Con el fin de lograr aumentar el número de ventas de tarjetas de crédito, se realiza una serie de estrategias como centrarse en tener mayor contactabilidad directa con el cliente a través de llamadas solo a celulares. Esto trae consigo una pérdida considerable de información y sobre todo de clientes potenciales.

Debido a que no se está tomando en cuenta las características del tipo de cliente al momento de ofrecer la tarjeta de crédito.

El presente trabajo de investigación tiene el objetivo de identificar el perfil de los clientes del departamento de Lima que aceptan una tarjeta de crédito por medio del canal de ventas de un Call Center.

Para ello se pretende utilizar el método del algoritmo CHAID Exhaustivo con el fin de identificar aquellos patrones que deriven en la aceptación de una tarjeta de crédito para así optimizar el número de llamadas, el manejo correcto de las bases de datos, el tiempo que se toma en concretar una venta vía telefónica, etc.



## **2.2. Formulación de las preguntas de investigación**

El problema de investigación es el siguiente:

¿Cuál es el perfil de los clientes del departamento de Lima de la base de datos enviada por el banco que aceptan una tarjeta de crédito vía Call Center.

## **2.3. Objetivos de la investigación**

Los objetivos del presente trabajo de investigación son los siguientes:

### Objetivo Principal:

Identificar el perfil de los clientes del departamento de Lima que aceptan una tarjeta de crédito por medio del canal de ventas de un Call Center.

### Objetivo secundario:

- Aplicar el algoritmo de Árboles de Clasificación CHAID Exhaustivo para identificar las variables independientes que determinan si el cliente tiene muchas posibilidades de aceptar una tarjeta de crédito.
- Medir la tasa de clasificación correcta del modelo.

## **2.4. Justificación de la investigación**

El manejo de la cartera de clientes potenciales que envía el banco se gestiona según las estrategias de ventas del supervisor; sin embargo al final de la gestión son medidas y pagadas por el banco. La entidad financiera por contrato está obligada a pagar comisiones por tarjeta de crédito activada en el mes.

El objetivo establecido por el banco es lograr el 25% de activaciones de tarjetas de crédito del total de la base entregada.

Según los últimos 5 meses de resultados de gestión se ha logrado llegar entre un 19% a 24% de activaciones de la base, lo que significa que las estrategias tomadas hasta ahora no han logrado pasar el 25%, es decir no se llega a meta establecida por la entidad financiera.

Por lo que se busca mejorar la estrategia para gestionar la cartera de clientes a través de la identificación del perfil de los clientes que aceptan una tarjeta de crédito vía Call Center y enfocarnos en primera instancia hacia a ellos, para poder ser más efectivos minimizando costos y tiempo en la gestión.

Reducción de costos, ya que disminuiría el número de llamadas, por ende disminuiría el recibo telefónico ya que el proveedor cobra S/.0,08centimos de sol el minuto.

Llamadas efectivas, al poder centrarse en los clientes con mayor potencial a adquirir una tarjeta de crédito, se reduciría el tiempo en la llamada para concretar la venta y optimizar el tiempo en lograr el objetivo de la gestión.

### **III. MARCO TEÓRICO**

#### **3.1. Árboles de Clasificación**

##### **3.1.1. Definición de Árboles de Clasificación**

Es uno de los métodos de aprendizaje no inductivo, supervisado y no paramétrico, muy utilizado. Destaca por su sencillez y puede utilizarse en diversas áreas como Marketing (segmentar clientes, identificar los perfiles de los clientes), en riesgo crediticio (determinación de alto riesgo, perfiles de grupos de riesgos basándose en la información del cliente), diagnósticos médicos, control de calidad, etc.

Según Berlanga Silvente, V., Rubio Hurtado, M. J., Vilà Baños, R. (2013) “Un árbol de Clasificación 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 encontrar subgrupos específicos y relaciones que tal vez no encontraríamos con estadísticos más tradicionales.

Los Árboles de Clasificación se definen como una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos, el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas.

Según Luque Martínez, T.(2000) los Árboles de Clasificación son “un conjunto de técnicas o algoritmos que, considerando una variable a explicar, persiguen la formación de grupos homogéneos, partiendo la población inicial secuencialmente respecto de esa variable que está explicada por otras variables independientes o explicativas”.

Los Árboles de Clasificación reubican criterios amplios con un centro de atención (nodo raíz) sobre los elementos importantes de una decisión, hacen resaltar premisas que con

frecuencia están escondidas y muestran el proceso de razonamiento mediante el cual se toman las decisiones bajo incertidumbre.

Según Schiattino I, y Silva C, (2008) los Árboles de Clasificación “permiten identificar perfiles o vectores que describen a un conjunto de individuos mediante una búsqueda que simplemente no es lineal sino más bien es comparable a una trayectoria de divisiones sucesivas: similar a una estructura arbórea. Los Árboles de Clasificación son flexibles, lo que permite identificar con relativa facilidad las combinaciones, nichos ecológicos o perfiles de predictores asociados a valores particulares de la respuesta de interés. El primer objetivo es predecir la clasificación que le correspondería a un sujeto con cierto perfil de valores en las variables explicativas; mientras que, en la segunda es estimar la respuesta de interés Y asociada a cada nicho”.

### 3.1.2. Elementos de los árboles de clasificación

Los Árboles de Clasificación presentan los siguientes elementos:

- **Nodo raíz o padre:** es el nodo base el cual se divide en subgrupos (dos o más según la técnica empleada) determinados por la partición de una variable predictora elegida, generando nodos hijos.
- **Nodos hijos o intermedios:** engendran dos o más (dependiendo del método empleado) segmentos descendientes inmediatos.
- **Nodos terminales:** es un nodo que no se puede dividir más, también denominado segmento terminal.
- **Rama de un nodo t:** consta de todos los segmentos descendientes del nodo t, excluyendo t. Es decir, los posibles caminos que se tienen de acuerdo a la decisión tomada.

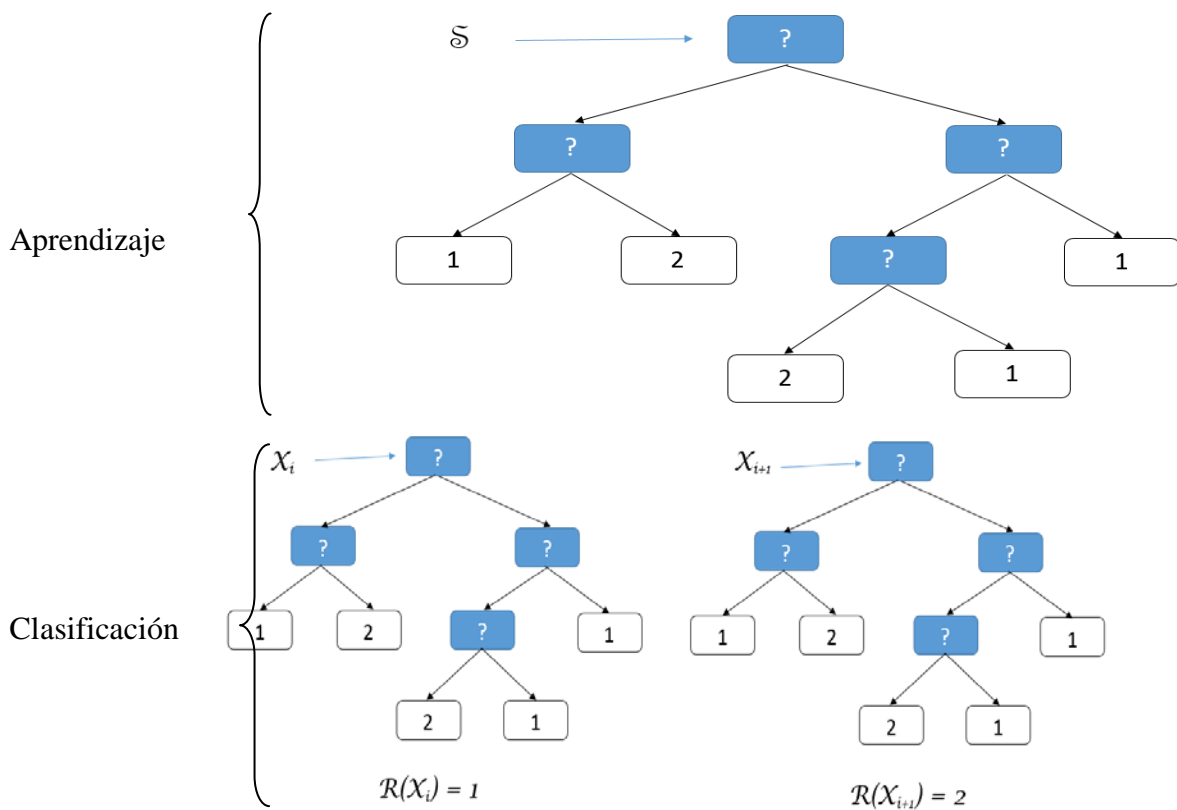
### 3.1.3. Fases de los árboles de clasificación

Los Árboles de Clasificación pueden ser descritos en dos fases:

- **Aprendizaje:** Consiste en la construcción del árbol a partir de un conjunto de prototipos(S). Constituye la fase más compleja y la que determina el resultado final. Dependerá de la decisión de construcción de este árbol el éxito de discriminación de variables que presenten relevancia en el estudio.

- Clasificación: Radica en el etiquetado de un patrón ( $X$ ), independiente del conjunto de aprendizaje. Se trata de responder a las preguntas asociadas a los nodos interiores utilizando los valores del atributo patrón ( $X$ ). Este proceso se repite desde el nodo raíz hasta alcanzar un nodo terminal, siguiendo el camino impuesto por el resultado de cada evaluación.

**Gráfico N° 1: Fases de Aprendizaje y Clasificación**



Fuente: Elaboración propia

### 3.1.4. Características de los Árboles de Clasificación

Para las características de un Árbol de Clasificación, Luque Martínez, T. (2000) menciona lo siguiente:

- “Hay un proceso iterativo de partición descendente.
- Surge la existencia de una variable criterio dependiente (monotética) o de varias variables (politéticas).
- Surge la existencia de variables independientes”.

Los Árboles de Clasificación se caracterizan por ser:

- **Segmentación:** Identifica a las personas que pueden ser miembros de un grupo específico.
- **Estratificación:** asigna casos a una categoría de entre varias.
- **Predicción:** crea reglas y son utilizadas para predecir eventos futuros.
- **Reducción de datos y clasificación de variables:** selecciona un subconjunto útil de variables predictoras a partir de una gran conjunto de variables para utilizarlo en la creación de un modelo paramétrico formal.
- **Identificación de interacción:** identifica las relaciones que pertenecen solo a subgrupos específicos para luego especificarlos en un modelo paramétrico formal.
- **Fusión de categorías y categorizar variables continuas:** recodifica las categorías de variables cualitativas y categoriza las variables continuas de un grupo de variables predictoras con una pérdida mínima de información. (Manual Decisión Trees 17.0)

### 3.1.5. Ventajas y desventajas de los Árboles de clasificación

#### Ventajas:

- Se obtiene conocimiento estructurado en forma de reglas de clasificación. Esto facilita interpretar en un lenguaje llano la caracterización de las clases o los valores de una variable de intervalo.
- Permite trabajar con todo tipo de variables predictoras: binarias, nominales, ordinales, de intervalo y de razón.
- Al tratarse de un procedimiento de análisis no paramétrico no se requiere validar supuestos distribucionales de probabilidad.
- Permite valores perdidos para las variables predictoras en los individuos, tanto en la fase de construcción del árbol como en la de predicción.
- Las reglas de asignación son simples y legibles, por tanto la interpretación de resultados es directa e intuitiva.
- Son bastante rápidos y la exigencia computacional no es muy alta.

### Desventajas:

- Las reglas de asignación son bastantes sensibles a pequeñas perturbaciones en los datos.
- Existe dificultad para elegir el árbol óptimo.
- Ausencia de una función global de las variables y como consecuencia pérdida de la representación.
- Los Árboles de Clasificación requieren un gran número de datos para asegurarse que la cantidad de las observaciones de los nodos hoja sea significativa.

### **3.1.6 Descripción de algoritmos de Árboles de Clasificación**

Existe una gran variedad de algoritmos que permiten la elaboración de un Árbol de Clasificación, los cuales difieren en las técnicas que utilizan respecto al tratamiento de los valores faltantes o la manera de podar los nodos.

Según el Manual del usuario SPSS DecisionTree17.0(2011)“Entre los algoritmos más conocidos se tienen:

- **CART.** Desarrollado por Breiman (1984), es un algoritmo de Árboles de Clasificación binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos. CART divide los datos en segmentos para que sean lo más homogéneos posible respecto a la variable dependiente.
- **CHAID.** Fue introducido por Kass (1980). 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.** Desarrollado por Biggs (1991), es un algoritmo completo para múltiples variables. 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.
- **QUEST.** Desarrollado por Loh y Shih (1997), es un algoritmo eficiente que crea un árbol binario sin sesgos con rapidez y precisión”.

## **3.2. Método CHAID Exhaustivo**

### **3.2.1 Definición del algoritmo CHAID Exhaustivo**

Es una modificación del algoritmo CHAID creado por Biggs & otros (1991) quienes proponen que el CHAID exhaustivo se debe enfocar en resolver los puntos débiles del CHAID. En concreto hay ocasiones en las que CHAID no encuentra la división óptima de una variable, ya que detiene la fusión de categorías en cuanto constata que todas las categorías restantes son distintas estadísticamente. Luego examina la serie de fusiones del predictor, busca el conjunto de categorías que proporciona la mayor asociación con la variable criterio y calcula un p-value corregido para esa asociación. De esta manera, el método CHAID Exhaustivo puede encontrar la mejor división para cada predictor y a continuación, elegir el predictor que se debe dividir comparando los niveles críticos corregidos.

El algoritmo utilizado en el CHAID Exhaustivo es muy similar al propuesto por Kass (1980). Funciona segmentando recursivamente cada agrupación o nodo no terminal en un conjunto de “k” nuevos nodos donde:  $2 \leq k \leq c$ , donde “c” es el número de categorías de variables predictoras.

Dado que el método de combinación de categorías de variables de CHAID Exhaustivo es más minucioso que el de CHAID, tarda más en realizar los cálculos. No obstante, cuando el tiempo no es un problema, el uso de CHAID Exhaustivo suele proporcionar más seguridad que CHAID.

### **3.2.2 Creación del algoritmo CHAID Exhaustivo**

Se puede resumir en 7 pasos importantes:

1. Para cada variable predictora X buscar la pareja de categorías de X cuya diferencia presenta la menor significación (es decir, la que presente el mayor nivel crítico) respecto a la variable criterio Y. El método usado para calcular el valor del nivel crítico depende del nivel de medida de Y.
  - Si Y es una variable continua, utilizar la prueba F de Fisher.



- Si Y es una variable nominal, se forma una tabla de contingencia con las categorías de X como filas y las categorías de Y como columnas. Utilizar la prueba chi-cuadrado de Pearson o la prueba de razón de verosimilitud.
  - Si Y es una variable ordinal, se ajusta un modelo de asociación de Y. Utilizar la prueba de razón de verosimilitud.
2. Unir en una categoría compuesta la pareja que proporcione el mayor nivel crítico.
  3. Calcular el nivel crítico basado en el nuevo conjunto de categorías de X. Recordar el nivel crítico y su correspondiente conjunto de categorías de X.
  4. Repetir los pasos 1, 2 y 3 hasta que solo queden dos categorías. Después, de entre todos los conjuntos de categorías de X, buscar aquel cuyo nivel crítico sea menor en el paso 3.
  5. Calcular el nivel crítico corregido mediante Bonferroni (es un test que evalúa la significación de las variables predictoras, proporciona un método de fijar un nivel de significación para la mejor segmentación encontrada de una determinada variable predictora, conservando el valor de primera especie, por debajo de cierto nivel crítico  $\alpha$  ) para el conjunto de categorías de X y las categorías de Y.
  6. Seleccionar la variable predictora X cuyo nivel crítico corregido sea el menor (aquella que sea la más significativa). Comparar su nivel crítico con el nivel alfa pre-establecido  $\alpha_{split}$  .
    - Si el nivel crítico es menor o igual que  $\alpha_{split}$  , dividir el nodo conforme al conjunto de categorías de X.
    - Si el nivel crítico es mayor que  $\alpha_{split}$  , no dividir el nodo. Este nodo es un nodo terminal.

7. Continuar con el proceso de desarrollo del árbol hasta que se cumpla una de las reglas de parada.

### 3.2.3 Precisión del algoritmo CHAID Exhaustivo

Después de generar un árbol, es muy importante señalar su precisión o validez predictiva. Si la variable criterio es de tipo nominal, la precisión indica el porcentaje de precisión correcta. Al contrario de esto, la imprecisión se llama *riesgo*. La estimación del riesgo y el error típico de la estimación del riesgo, son valores que indican la capacidad del indicador.

Se puede estimar el riesgo del árbol mediante tres métodos distintos:

- La reestimación de toda la muestra.
- Validación por división muestral.
- Validación cruzada.

El método más sencillo para calcular el riesgo es el de la reestimación, pero normalmente subestima el riesgo verdadero. La validación por división muestral, es un buen método cuando el conjunto de datos es suficientemente grande. El riesgo se calcula a partir de la muestra de comprobación. La validación cruzada implica dividir la muestra en una serie de muestras más pequeñas y se calcula como el promedio de todos los árboles generados.

### 3.2.4 Ventajas y desventajas del método

#### Ventajas:

**Transparencia:** La forma de un árbol CHAID exhaustivo es intuitiva. Esto significa que el usuario puede comprobar la racionalidad del modelo y, si lo cree necesario, modificar el árbol en su arquitectura según su propia experiencia o conocimiento.

**Construcción sencilla:** Los modelos CHAID exhaustivo pueden manejar fácilmente variables independientes tanto categóricas (como el estado civil) como continuas relacionadas (como los ingresos). Concretamente, si las variables independientes son categóricas y con un alto nivel de cardinalidad (que implícitamente "contienen" relaciones) CHAID exhaustivo dará sin dudas un mejor resultado.

**Seguridad:** El CHAID exhaustivo suele proporcionar más seguridad, debido a que su método de combinación de categorías de variables es más minucioso que el análisis CHAID.

### **Desventajas**

**El volumen de datos:** Requiere grandes masas de datos para asegurarse de que la cantidad de observaciones de los nodos hijos sea significativa.

**Las variables continuas han de estar relacionadas:** Las variables independientes continuas, como los ingresos, deben estar relacionadas en grupos a imagen de los datos categóricos antes de ser usadas en CHAID exhaustivo.

**Tiempo:** Tarda más tiempo en realizar los cálculos de las divisiones de los nodos porque realiza un análisis más minucioso.

## **IV. METODOLOGÍA DE LA INVESTIGACIÓN**

### **4.1. Tipo de investigación**

El tipo de investigación utilizado es no experimental y es de carácter exploratorio. Según Roberto Hernández Sampieri, Carlos Fernández Collado y Pilar Baptista Lucio. (2003) Un estudio es exploratorio “cuando el objetivo es examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes. Es decir cuando la revisión literaria reveló que tan solo hay guías no investigadas e ideas vagamente relacionadas con el problema de estudio, o bien, si deseamos indagar sobre temas y áreas desde nuevas perspectivas”.

### **4.2. Formulación de la Hipótesis**

La hipótesis de la investigación es la siguiente:

A través del algoritmo CHAID Exhaustivo se puede construir un patrón de comportamiento de los clientes identificando las variables más significativas que puedan explicar su decisión de aceptar una tarjeta de crédito vía Call Center.

### **4.3. Identificación de variables**

Se definen las siguientes variables:

- **Variable dependiente (Y)**

Esta variable surge del último estado registrado por los agentes del Call Center en el sistema.

En la presente investigación se utilizó la variable ACEPTACIÓN como variable respuesta, la cual toma dos valores:

1 :Aceptó TC

0 :No aceptó TC

El presente trabajo se enfoca en dos tipos de resultados: Cliente aceptó TC y Cliente no aceptó TC. Este último tipo combina los resultados que han sido generados de cada una de las llamadas como son: contacto con terceros, cliente no desea producto, no contacto, etc.

**Cliente aceptó TC**, aquellos clientes que a través del contacto directo con el titular vía llamada telefónica acepta una tarjeta de crédito.

Esta calificación se codificará con el valor “0” y etiquetada como “ACEPTO TC”.

**Cliente no aceptó TC**, aquellos clientes que por diversas razones no adquirieron una tarjeta de crédito durante la campaña actual.

Esta calificación será codificada con valor “1” y etiquetada como “NO ACEPTO TC”.

- **Variables independientes:**

- Campaña.- Es el segmento al que asigna el banco a cada persona.

- 1 : Club Sueldo

- 2 : Convenios

- 3 : Dependientes A

- 4 : Dependientes B

- 5 : Independientes

- Tipo de Cliente.- Representa al grupo de clientes al que es segmentado la persona por parte del banco.

- 1 : Cliente

- 2 : Ex Cliente

- 3 : No Cliente

- Ingreso Neto.- Es el monto en soles (S/.) resultante de restarle al monto Bruto de la persona todas las deducciones que le corresponde.

- Edad.- Son los años cumplidos que tiene el cliente a la hora de ofrecerle una tarjeta de crédito.

- Género del Cliente.-

1 : Femenino

2 : Masculino

- Estado civil.- Según las leyes peruanas y por división del banco se clasifican en 4 grupos.

1 : Soltero

2 : Casado

3 : Viudo

4 : Divorciado

- Tipo de Tarjeta de Crédito.- Es el tipo de tarjeta que con análisis previos y propios del banco se le ofrece al cliente.

1 : MC BLACK

2 : MC ORO

3 : MC PLATINUM

4 : VISA CLASICA

5 : VISA ORO

6 : VISA ORO LIGHT

7 : VISA PLATINUM

8 : VISA SIGNATURE

- Límite de Tarjeta de Crédito.- Muestra el límite de crédito establecido por la empresa para el cliente (en soles).

- Cliente Bancarizado.- Muestra el grupo al que pertenece el cliente según su experiencia bancaria.

1 :Bancarizado (Cliente con experiencia bancaria)

2 : No Bancarizado (Cliente sin experiencia bancaria)

La descripción del tipo de variable y codificación se detalladaa continuación:

**Cuadro N° 1: Descripción de las variables**

	VARIABLE	TIPO DE VARIABLE	VALORES	ETIQUETA
<b>Dependiente</b>	Aceptación	Nominal	1	Aceptó TC
			0	No aceptó TC
<b>Independientes</b>	Campaña	Nominal	1	ClubSueldo
			2	Convenios
			3	Dependientes A
			4	Dependientes B
			5	Independientes
	Tipo_Cliente	Nominal	1	Cliente
			2	Ex Cliente
	Género del Cliente	Nominal	3	No Cliente
			0	Femenino
	Estado Civil	Nominal	1	Masculino
2			Soltero	
3			Casado	
4			Viudo	
Tipo de Tarjeta	Nominal	1	Divorciado	
		2	MC BLACK	
		3	MC ORO	
		4	MC PLATINUM	
		5	VISA CLASICA	
		6	VISA ORO	
		7	VISA ORO LIGHT	
		8	VISA PLATINUM	
Cliente Bancarizado	Nominal	1	VISA SIGNATURE	
		2	Bancarizado	
	Edad	Cuantitativa continua		No Bancarizado
	Límite_TC	Cuantitativa continua		
	Ingreso_Neto	Cuantitativa continua		

Fuente: Elaboración propia.

#### 4.4. Población y muestra

La población está conformada por toda la base de datos de clientes recibida los últimos 5 meses en que fue implementado el nuevo sistema de ACD GoautoDial (Diciembre 2013 – Abril 2014) conformada por 160,479 clientes pertenecientes al departamento de Lima.

Para la muestra se considera una población  $N = 160,479$  clientes,  $p = 0.5$ , un nivel de confianza del 98%, un error  $e = 0.005$  y se obtiene que:

$$n = \frac{N \times Z_{1-\frac{\alpha}{2}}^2 \times p \times q}{e^2 \times (N - 1) + Z_{1-\frac{\alpha}{2}}^2 \times p \times q} = 40,471$$

La muestra cuenta con un total de 40,471 clientes que se utilizará para trabajar con el modelo del algoritmo CHAID Exhaustivo.

#### 4.5. Instrumento de colecta de datos

La herramienta empleada para la obtención de las ventas de tarjetas de crédito es el Software de Monitoreo ACD GoAutoDial 3.0 el cual se encuentra instalado en la Unidad de Negocios de la Entidad Financiera dentro del Call Center. En ella se registran todas las llamadas que se realizan a los clientes sean llamadas contestadas, no contestadas, perdidas, cortadas, etc.



## V. PROCEDIMIENTO DE ANÁLISIS DE DATOS

Para todo el procedimiento de análisis de datos del presente estudio se utilizará el programa SPSS Statistics 22.0.

Para poder identificar el perfil de los clientes que aceptan una tarjeta de crédito se realizará un análisis descriptivo de cada una de las variables con el fin de conocer cómo se comportan individualmente y como se asocian con la variable de aceptación de una tarjeta crédito.

Luego se analizará el modelo del algoritmo CHAID Exhaustivo con el fin de identificar el perfil de los clientes que aceptan una tarjeta de crédito en el departamento de Lima.

El procedimiento de análisis de datos que se aplicará es el siguiente:

**Paso 1:** Realizar un análisis univariado y bivariado de la variable dependiente (Y: Aceptación) y de las variables independientes ( $X_i$ ). Utilizando las tablas de frecuencia y tablas cruzadas para ver cómo están distribuidas las frecuencias presentándolas en términos porcentuales.

**Paso 2:** Mostrar los resultados del Árbol de Clasificación CHAID Exhaustivo cuyos parámetros para la obtención del mismo fueron los siguientes:

- Método de validación: Muestra de división
- Profundidad del árbol: 3.
- Tamaño del nodo parental: 400.
- Tamaño del nodo filial: 200.

**Paso 3:** Realizar el Análisis de Ganancias para los nodos, el cual nos ofrece un resumen de información sobre los nodos terminales del modelo.

**Paso 4:** Realizar el análisis de la estimación del riesgo de clasificar erróneamente a un cliente dentro de la variable respuesta pronosticada.

**Paso 5:** Analizar la tabla de clasificación de los clientes, que proporciona una rápida evaluación de la bondad del funcionamiento del modelo; es decir, esta tabla muestra el número de casos clasificados correcta e incorrectamente para cada categoría de la variable dependiente.

**Paso 6:** Determinar el modelo que mejor identifiquen el perfil de los clientes que aceptan una tarjeta vía Call Center.

## VI. RESULTADOS

### 6.1. Análisis Univariado y Bivariado

Empezamos examinando los efectos de las variables independientes sobre la única variable ACEPTACIÓN una a una de manera independiente:

**Cuadro N°2: Tabla de frecuencias de las variables cualitativas**

		Porcentaje (%)
ACEPTACION (Y)	NO ACEPTO TC	76,8
	ACEPTO TC	23,2
	Total	100,0
CAMPAÑA (X1)	Club Sueldo	,8
	Convenios	7,5
	Dependientes A	19,7
	Dependientes B	10,6
	Independientes	61,4
	Total	100,0
TIPO_CLIENTE (X2)	Cliente	12,0
	Ex Cliente	22,6
	No Cliente	65,4
	Total	100,0
GENERO_CLIENTE (X5)	Femenino	46,9
	Masculino	53,1
	Total	100,0
ESTADO_CIVIL (X6)	Soltero	37,1
	Casado	38,6
	Viudo	7,5
	Divorciado	16,8
	Total	100,0
TIPO_TARJETA (X7)	MC BLACK	,0
	MC ORO	,5
	MC PLATINUM	,2
	VISA CLASICA	61,5
	VISA ORO	2,9
	VISA ORO LIGHT	30,8
	VISA PLATINUM	3,8
	VISA SIGNATURE	,3
	Total	100,0
CLIENTE_BANCARIZADO (X9)	Bancarizado	99,9
	No Bancarizado	,1
	Total	100,0

Fuente: Elaboración propia

Podemos observar que del total de clientes, el 23,2% aceptó una tarjeta a través del canal de ventas Call Center, mientras que 76.8% no aceptó la tarjeta.

Los clientes que trabajan de forma independiente conforman el 61.4% de la base de datos, y las personas que no son clientes de la entidad financiera conforman el 65.4%.

El 53.1% de la base de datos está conformado por varones. La tarjeta que más se ofrece es la VISA CLÁSICA con una participación de 61.5% seguida de la VISA ORO LIGHT con un 30.8%., finalmente el 99.9% de la base de datos está conformado por clientes Bancarizados, es decir, clientes que tienen experiencia en el rubro financiero.

**Cuadro N°3: Tabla de Contingencias de las variables.**

		ACEPTACION	
		ACEPTO TC	NO ACEPTO TC
<b>CAMPAÑA</b>	Club Sueldo	25,7%	74,3%
	Convenios	22,3%	77,7%
	Dependientes A	24,6%	75,4%
	Dependientes B	24,2%	75,8%
	Independientes	22,6%	77,4%
<b>TIPO_CLIENTE</b>	Cliente	24,0%	76,0%
	Ex Cliente	22,7%	77,3%
	No Cliente	23,2%	76,8%
<b>GENERO_CLIENTE</b>	Femenino	21,7%	78,3%
	Masculino	24,4%	75,6%
<b>ESTADO_CIVIL</b>	Soltero	25,0%	75,0%
	Casado	23,1%	76,9%
	Viudo	15,6%	84,4%
	Divorciado	22,8%	77,2%
<b>TIPO_TARJETA</b>	MC BLACK	50,0%	50,0%
	MC ORO	21,5%	78,5%
	MC PLATINUM	22,9%	77,1%
	VISA CLASICA	22,8%	77,2%
	VISA ORO	27,1%	72,9%
	VISA ORO LIGHT	23,7%	76,3%
	VISA PLATINUM	21,7%	78,3%
	VISA SIGNATURE	19,1%	80,9%
<b>CLIENTE_BANCARIZADO</b>	Bancarizado	23,2%	76,8%
	No Bancarizado	0,0%	100,0%

*Fuente: Elaboración propia*

Los clientes de *ClubSueldos* son los que tienen mayor propensión a aceptar una tarjeta de crédito con un 25.7%, seguido de los clientes dependientes A.

Para la variable TIPO\_CLIENTE, el 24% aceptan una tarjeta de crédito. Se puede observar que el 24.4% de los varones aceptan una tarjeta de crédito, superando a las mujeres y que el 25% de clientes solteros es el grupo más representativo en aceptar una tarjeta de crédito, seguido por los clientes casados con un 23.1%.

Con respecto al tipo de tarjeta que aceptan los clientes en mayor propensión es la VISA CLASICA con un 27.1% seguida de la VISA ORO LIGHT con un 23.7% y que del total de clientes bancarizados el 23.2% aceptan una tarjeta de crédito vía Call Center.

## 6.2. Árbol de Clasificación CHAID Exhaustivo

Se consideró trabajar con un conjunto de prueba que es usada para la determinación del modelo, es decir, el conjunto de datos original se dividió en dos grupos, uno de entrenamiento y otro de prueba.

Con la validación por división muestral, se genera el modelo utilizando una muestra de entrenamiento y después se pone a prueba con una muestra de reserva, este método de validación funciona bien con archivos de datos grandes.

Para este modelo se consideró el 80% de la muestra para entrenamiento y el 20% de prueba, basado en la lógica de Pareto, que quiere decir que el 20% de los clientes explican el 80% del comportamiento de la población.

**Cuadro N° 4: Tabla de resumen del modelo**

Resumen del modelo		
Especificaciones	Método de crecimiento	EXHAUSTIVE CHAID
	Variable dependiente	ACEPTACION
	Variables independientes	CAMPAÑA, TIPO_CLIENTE, INGRESO_NETO, EDAD, GENERO_CLIENTE, ESTADO_CIVIL, TIPO_TARJETA, LIMITE_TC, CLIENTE_BANCARIZADO
	Validación	Muestra de división
Resultados	Máxima profundidad del árbol	3
	Casos mínimos en nodo padre	400
	Casos mínimos en nodo hijo	200
	Variables independientes incluidas	EDAD, TIPO_CLIENTE, GENERO_CLIENTE, CAMPAÑA, ESTADO_CIVIL, TIPO_TARJETA, INGRESO_NETO
	Número de nodos	62
	Número de nodos terminales	39
	Profundidad	3

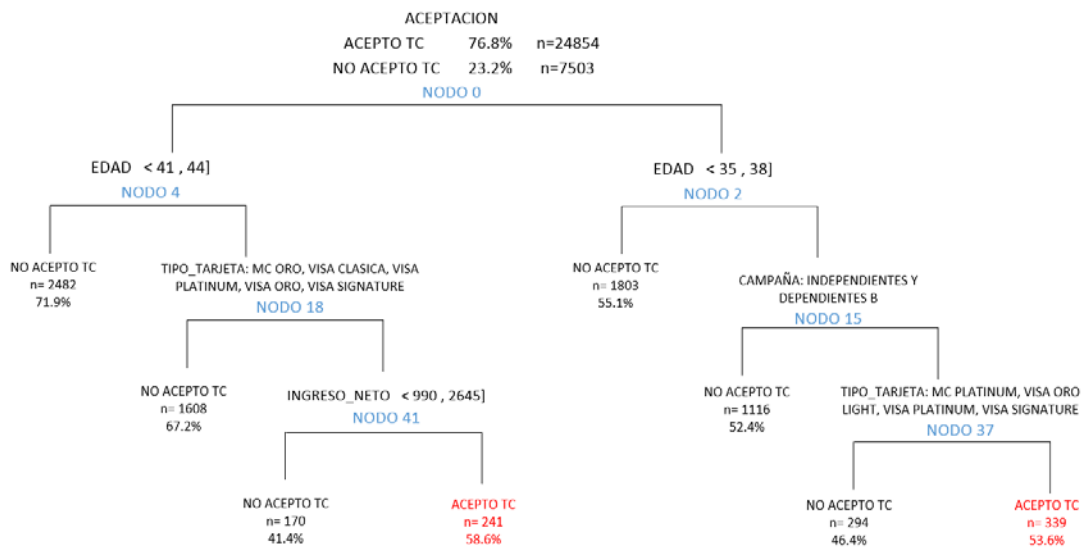
Fuente: SPSS Statistics 22.0

Como podemos observar en el Cuadro N° 4, se especificó el método a utilizar que es el de CHAID Exhaustivo, se utilizaron las nueve variables predictoras y que el método de validación que se utilizó fue el de *Muestra de división*, donde internamente se tomó una muestra de entrenamiento del 80% y como muestra de comprobación el 20% de los datos.

Como se mencionó, se utilizaron nueve variables independientes, pero solo se han incluido siete en el modelo final. La variable *Cliente\_Bancarizado y Límite de TC* no contribuyen de forma significativa al modelo, por lo que se eliminaron automáticamente del modelo final.

Los resultados obtenidos al utilizar el método CHAID Exhaustivo muestran un árbol con un total de 3 niveles, 62 nodos y 39 nodos terminales

**Gráfico N° 2: Árbol de Clasificación – Método CHAID EXHAUSTIVO**



Fuente: Elaboración propia

Al utilizar el método de CHAID Exhaustivo, podemos ver que la variable EDAD es el mejor predictor para la Aceptación de una tarjeta de crédito, ya que su p-valor es significativo, es decir menor a 0.05 y el valor de su  $X^2$  calculado es el de mayor valor.

**Cuadro N° 5: Ganancias para los nodos: ACEPTA TC**

	Nodo	Nodo		Ganancia		Respuesta	Índice
		N	Porcentaje	N	Porcentaje		
Entrenamiento	<b>41</b>	<b>411</b>	<b>1,3%</b>	<b>241</b>	<b>3,2%</b>	<b>58,6%</b>	<b>252,9%</b>
	<b>37</b>	<b>633</b>	<b>2,0%</b>	<b>339</b>	<b>4,5%</b>	<b>53,6%</b>	<b>231,0%</b>
	36	1495	4,6%	673	9,0%	45,0%	194,1%
	35	748	2,3%	333	4,4%	44,5%	192,0%
	13	1786	5,5%	758	10,1%	42,4%	183,0%
	33	532	1,6%	200	2,7%	37,6%	162,1%
	34	399	1,2%	127	1,7%	31,8%	137,3%
	39	834	2,6%	257	3,4%	30,8%	132,9%
	42	224	0,7%	69	0,9%	30,8%	132,8%
	48	291	0,9%	86	1,1%	29,6%	127,4%
	32	273	0,8%	79	1,1%	28,9%	124,8%
	11	716	2,2%	202	2,7%	28,2%	121,7%
	40	1757	5,4%	474	6,3%	27,0%	116,3%
	58	730	2,3%	185	2,5%	25,3%	109,3%
	38	1337	4,1%	315	4,2%	23,6%	101,6%
	24	974	3,0%	229	3,1%	23,5%	101,4%
	47	1346	4,2%	313	4,2%	23,3%	100,3%
	53	952	2,9%	214	2,9%	22,5%	96,9%
	43	440	1,4%	98	1,3%	22,3%	96,1%
	17	1240	3,8%	247	3,3%	19,9%	85,9%
	55	420	1,3%	82	1,1%	19,5%	84,2%
	31	421	1,3%	82	1,1%	19,5%	84,0%
	23	1988	6,1%	355	4,7%	17,9%	77,0%
	50	1121	3,5%	179	2,4%	16,0%	68,9%
	54	1336	4,1%	204	2,7%	15,3%	65,9%
	61	1960	6,1%	290	3,9%	14,8%	63,8%
	44	620	1,9%	88	1,2%	14,2%	61,2%
	52	436	1,3%	61	0,8%	14,0%	60,3%
	46	311	1,0%	40	0,5%	12,9%	55,5%
	22	606	1,9%	77	1,0%	12,7%	54,8%
57	211	0,7%	25	0,3%	11,8%	51,1%	
59	238	0,7%	28	0,4%	11,8%	50,7%	
9	3403	10,5%	370	4,9%	10,9%	46,9%	
60	507	1,6%	52	0,7%	10,3%	44,2%	
30	484	1,5%	44	0,6%	9,1%	39,2%	
51	255	0,8%	23	0,3%	9,0%	38,9%	
45	349	1,1%	27	0,4%	7,7%	33,4%	
56	343	1,1%	23	0,3%	6,7%	28,9%	
49	230	0,7%	14	0,2%	6,1%	26,3%	

Método de crecimiento: EXHAUSTIVE CHAID

Variable dependiente: ACEPTACION

Fuente: SPSS Statistics 22.0

En el Cuadro N° 5, podemos observar las ganancias para los nodos, el cual nos ofrece un resumen de información sobre los nodos terminales del modelo. Con frecuencia, el único interés lo suscitan los nodos terminales, ya que representan los mejores pronósticos de clasificación para el modelo.

Podemos observar que los nodos 41, 37, 36, 35, 13, 33, 34, 39, 42, 48, 32, 11, 40, 58, 38, 24 y 47 tienen un índice mayor al 100% siendo los nodos 41 y 37 los que mayor probabilidad de aceptar una tarjeta de crédito dentro de sus patrones

**Cuadro N° 6: Nodos con mayor índice de Ganancias**

NODO 41			NODO 37		
CATEGORIA	n	%	CATEGORIA	n	%
ACEPTO TC	241	58,64%	ACEPTO TC	339	53,55%
NO ACEPTO TC	170	41,36%	NO ACEPTO TC	294	46,45%

Fuente: Elaboración propia

El nodo 41 nos indica que el 58.6% de clientes que tienen un ingreso neto entre S/.990 soles y S/.2,645 soles, cuya edad oscila entre los 41 y 44 años tienden a aceptar una tarjeta de crédito que puede ser una MC Oro, Visa Oro, Visa Clásica, Visa Platinum o Visa Signature.

Mientras que el Nodo 37 muestra que el 53.5% de los clientes que aceptan una tarjeta de crédito son trabajadores independientes y dependientes cuya edad oscila entre los 35 y 38 años y el tipo de tarjeta que acepta es MC Platinum, Visa Oro Light, Visa Platinum o Visa Signature.

**Cuadro N° 7: Tabla de riesgos del modelo**

Riesgo		
Ejemplo	Estimación	Error estándar
Entrenamiento	,108	,002
Prueba	,077	,005

Método de crecimiento: EXHAUSTIVE CHAID

Variable dependiente: ACEPTACION



La estimación del riesgo es de 0.108, lo cual indica que la categoría pronosticada por el modelo (Aceptación o no de una tarjeta de crédito vía Call Center) es errónea para el 10.8% de los casos. Esto quiere decir entonces que el riesgo de clasificar erróneamente a un cliente es aproximadamente el 10.8%.

**Cuadro N° 8: Tabla de Clasificación correcta del modelo**

		Clasificación		
		Pronosticado		
Ejemplo	Observado	NO ACEPTA TC	ACEPTA TC	Porcentaje correcto
Entrenamiento	NO ACEPTA TC	22420	435	98,1%
	ACEPTA TC	3075	6447	67,7%
	Porcentaje global	78.7%	21.3%	89.2%
Prueba	NO ACEPTA TC	6115	123	98,0%
	ACEPTA TC	498	1358	66,8%
	Porcentaje global	81.7%	18.3%	92.3%

Método de crecimiento: EXHAUSTIVE CHAID

Variable dependiente: ACEPTACION

Fuente: SPSS Statistics 22.0

El tipo de validación del modelo que se utilizó fue la validación de división muestral, y como se puede observar, la muestra de comprobación que fue el 20% de los datos demuestra que el 92.3% de los clientes están correctamente clasificados por lo que se comprueba que el modelo clasifica correctamente a los clientes.

La tabla de clasificación nos muestra un problema potencial con este modelo. Para aquellos clientes que aceptan una tarjeta de crédito vía Call Center, pronostica una valoración positiva para solo 67.7% de ellos, lo que significa que el 32.3% de los clientes con una valoración de aceptación de tarjeta de crédito positiva aparecen inapropiadamente clasificados como clientes que no aceptarían la tarjeta de crédito.

### 6.3. Comparación entre el modelo elegido y resultados reales

Del Árbol de Clasificación final obtenido se seleccionan los dos perfiles encontrados que representan la mayor probabilidad de aceptación de tarjetas de crédito viaCall Center de clientes en el departamento de Lima.

El cuadro N° 9 se muestra a continuación los resultados reales sin aplicar el modelo elegido.

**Cuadro N° 9: Escenario real de aceptación de tarjetas de crédito**

MES	CANTIDAD DE CLIENTES	NIVEL DE ACEPTACIÓN %
<b>dic-13</b>	10.682	23,48%
<b>ene-14</b>	11.119	22,10%
<b>feb-14</b>	6.487	23,76%
<b>mar-14</b>	9.180	24,95%
<b>abr-14</b>	3.003	19,40%

Fuente: Elaboración propia.

Se puede observar que el nivel de aceptación de tarjetas de crédito en los últimos cinco meses ha oscilado entre un 19.40% y 24.95% como máximo, no pudiendo llegar a la fecha al 25% de objetivo que establece la entidad financiera.

**Cuadro N° 10: Escenario de pronóstico de aceptación de tarjetas de crédito**

**Usando el modelo**

MES	CANTIDAD DE CLIENTES	NIVEL DE ACEPTACIÓN %
<b>dic-13</b>	2.136	38,35%
<b>ene-14</b>	2.224	36,09%
<b>feb-14</b>	1.297	38,81%
<b>mar-14</b>	1.836	40,74%
<b>abr-14</b>	601	31,68%

Fuente: Elaboración propia.

Según se muestra en el cuadro N°10, considerando el modelo obtenido se logra aumentar el nivel de aceptación de tarjetas de crédito

## VII. CONCLUSIONES

- En base a los resultados mostrados se identificó dos perfiles de clientes en departamento de Lima que aceptan una tarjeta de crédito vía Call Center:

*“Clientes cuya edad oscila entre los 41 y 44 años que tienen un ingreso neto entre S/.990 soles a S/.2,645 soles y tienden a aceptar una tarjeta de crédito que puede ser una MC Oro, Visa Oro, Visa Clásica, Visa Platinum o Visa Signature”.*

*“Clientes cuya edad oscila entre los 35 y 38 años, siendo trabajadores independientes o dependientes B que aceptan una tarjeta de crédito de tipo MC Platinum, Visa Oro Light, Visa Platinum o Visa Signature”.*

- Podemos concluir entonces que las variables que mejor se asocian o explican el modelo de aceptación de una tarjeta de crédito vía Call Center son las variables: edad, ingresos netos mensuales, el tipo de trabajador(variable Campaña) y el tipo de tarjeta que se le ofrece al cliente.
- Al aplicar este modelo a la estrategia de ventas se puede concluir que ayuda al incremento del indicador del nivel de aceptación, mejorando la gestión, minimizando tiempo y costos.

## VIII. RECOMENDACIONES

- Realizar previamente una limpieza de datos.
- Utilizar este algoritmo como un modelo inicial de estrategia de captación de nuevos clientes potenciales de Lima.
- Se recomienda utilizar otros algoritmos de Árboles de Clasificación como el CART, C5, Quest, con el fin de comparar resultados y ver si estos últimos aportan mejoras en el modelo de aceptación de tarjetas de crédito.
- Establecer un sistema de información en línea que permita identificar a los clientes de la base de datos que entrega la entidad financiera cuyo perfil se adecue al modelo planteado de manera que mande alertas a los agentes del Call Center antes de hacer las llamadas.
- El modelo debe de seguir actualizándose con la nueva información disponible.

## IX. BIBLIOGRAFÍA

- 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>
- Breiman, L. 1984. Classification and Regression Trees. California, US.
- Hernández, O., José, Ramírez, Q., José, Ferrari, R., Introducción a la minería de datos. Madrid, España. Editorial Pearson Prentice Hall, última reimpresión, 2007. 656p.
- Hernández Sampieri Roberto, Fernández Collado Carlos y Baptista Lucio Pilar. (2003). Metodología de la investigación. Mexico, DF. McGraw-Hill Interamericana. Tercera Edición.
- IBM SPSS (Statistical Package for the Social Sciences Inc, US). 2011. Manual del usuario SPSS DecisionTree 17.0
- Luque Martínez, T. 2000. Técnicas de análisis de datos en investigación de mercados. Madrid, ES. Ediciones Pirámide Grupo Anaya S.A.
- Pérez López, César y Santin González, D. Minería de datos, técnicas y herramientas, España, editorial Thomson, 2007. 789p.
- SPSS (Statistical Package for the Social Sciences Inc, US). 1998. Manual del usuario del Answer Tree 2.0.

## **X. ANEXO**

# Anexo N° 1: Árbol de clasificación utilizando el algoritmo CHAID Exhaustivo

