

Modelo de Árboles de Clasificación para definir el perfil de un alumno leal de la Facultad de Ingeniería y Arquitectura de la Universidad Peruana Unión

Millones Liza, Jhonny¹; Barrantes Delgado, Jhonatan²; Vivas Quispe, Alexander
Dr. Palza Vargas, Edgardo; Dr. Mamani Apaza, Guillermo

Resumen

El presente trabajo de investigación es determinar el perfil de lealtad de los alumnos de la Facultad de Ingeniería y Arquitectura, para el cual se ha definido el concepto de lealtad como: “el compromiso que se tiene hacia una marca o servicio”, luego se identifican los tipos de lealtad que existen y así podremos definir el perfil del alumno leal, esta será una herramienta útil en la mitigación de falencias educativas que se presentan a menudo, como son: el retiro de alumnos, la insatisfacción, la poca participación en las actividades. Se piensa trazar o diseñar una ruta de implementación la cual será a corto, mediano y largo plazo, este brindara el camino a seguir para alcanzar mejores prácticas consiguiendo mejorar el nivel de lealtad de los alumnos. Todo esto haciendo uso de Árboles de Clasificación (AC).

Para el caso de proyectos de implementación de minería de datos, utilizaremos la metodología CRISP-DM. CRISP-DM es una organización europea creada por tres grandes empresas especialistas en proyectos de minería de datos que son SPSS, NCR y DaimlerChrysler. Esta metodología permite desarrollar proyectos de minería de datos bajo un proceso estandarizado de definición y validación, esto ayuda a que el costo sea razonable y con un alto impacto en el negocio. Se incluye un modelo y una guía, estructurada en seis fases, las cuales son: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.

Para recopilar la información, se utilizó un cuestionario de lealtad y para identificar los atributos del perfil del alumno leal se aplicó el modelo de Árboles de Clasificación.

El cuestionario se aplico a una muestra de 200 alumnos, de una población de 541 que forman parte de la FIA.

El modelo de Arboles de Clasificación nos ha permitido identificar los perfiles de los cuatro tipos de alumnos leales: terrorista, rehén, mercenario, apóstol.

1. Introducción

El presente trabajo de investigación tiene como objeto de estudio la lealtad o fidelidad del alumno. Según Seto en su libro “De la calidad del servicio a la fidelidad del cliente” afirma que los beneficios de tener clientes fieles generan un incremento en los ingresos de la empresa, esto se origina porque los clientes generan recomendaciones boca-oído hacia otros posibles clientes. **Ayudan a disminuir los costes ya que es menos costoso de servir.**

¹E.A.P. de Ingeniería de Sistemas, Universidad Peruana Unión. Jhonny859@hotmail.com

²E.A.P. de Ingeniería de Sistemas, Universidad Peruana Unión. jhonatanbarrantes@hotmail.com

Por otro lado Vladilo en su libro “Lealtad a grandes rasgos” nos muestra que “*La competitividad de las organizaciones depende hoy en día de su capacidad para captar y FIDELIZAR*”, por ende los clientes leales siempre han sido importantes para el éxito a largo plazo de una empresa. Siendo que los clientes leales son parte importante en la mejora de una organización, en el caso de la FIA, obtener y comprender el perfil de lealtad del alumno, le permitirá adaptar y crear mejores estrategias para obtener la atención y lograr que el alumno se sienta identificado con la Facultad. Ubicaremos a los alumnos en cuatro niveles de lealtad: Apóstoles, Rehenedos, Mercenarios y

Terroristas. De aquí nace la necesidad de crear un modelo de árboles de Clasificación haciendo uso de la herramienta Clementine, esta agrupara los datos de la muestra a partir de los atributos más importantes, así obtendremos el perfil de un alumno leal de la FIA **para de esa manera generar escenarios, pronósticos y reportes que apoyan a la toma de decisiones.** Todo esto se traduce como una ventaja competitiva al negocio. Para la presente investigación utilizamos la metodología CRISP-DM porque es la que más se ajusta a los proyectos realizados con minería de datos, esta metodología se divide en 6 fases, las cuales usamos como guía para el desarrollo de nuestra investigación:

Comprensión del negocio: En esta primera fase se determina los objetivos de la Facultad de Ingeniería y Arquitectura. En particular se evalúa la situación en que se encuentra la FIA encontrando los puntos críticos a ser mejorados.

Comprensión de los datos: En esta segunda fase realizamos una recopilación inicial de los datos, se tiene una muestra de 200 alumnos obtenida mediante encuestas de lealtad (estudio no paramétrico), luego se hizo una descripción y exploración de los datos para verificar la calidad de la información.

Preparación de los datos: En esta fase seleccionamos, depuramos, construimos e integramos los datos para posteriormente analizarlos en la herramienta de trabajo (Clementine).

Modelado: En esta fase comenzamos con la construcción y evaluación del modelo el cual es Árboles de Clasificación.

Evaluación: En esta fase pasamos a evaluar los resultados obtenidos del modelo desarrollado en la herramienta Clementine.

Despliegue: En esta última fase realizamos una revisión general del proyecto para de esa manera corregir algún posible error que se pueda encontrar.



Figura 1. Diagrama de la metodología CRISP-DM

2. Que es Perfil

Se entiende como un conjunto de características que tiene un elemento, cada cliente es distinto. Comprender los objetivos explícitos e intuir los objetivos implícitos del cliente permitirá al gestor fijar las bases de una buena relación personal y un servicio a medida que puedan ser el sustento de una relación duradera y mutuamente provechosa. Esos objetivos son la herramienta para definir las líneas estratégicas de inversión que, permitirán obtener un binomio rentabilidad-riesgo adecuado para el cliente. Es por eso que cuando nos referimos al perfil de lealtad del alumno los distribuimos en 4 categorías con una serie de rasgos comunes que expresen un tipo de compromiso con la FIA.

3. Qué es Lealtad

Según (Oliver 1997) expresa que “La lealtad es un compromiso profundo de recomprar en el futuro un producto/servicio preferido, generando así compras de repetición de la misma marca o conjunto de marcas, a pesar de que las influencias situacionales y los esfuerzos de marketing tengan

el potencial de provocar un comportamiento de cambio”.

Lo divide en cuatro Fases:

- A) *Lealtad cognitiva*: se da la información disponible para el consumidor sobre los atributos de una marca, indica que una marca es preferida a sus alternativas.
- B) *Lealtad afectiva*: está asociada a las sensaciones o sentimientos respecto a la marca. se desarrolla un vínculo o actitud, en base a una acumulación de sus satisfactorios.
- C) *Lealtad conativa*: influenciado por episodios repetitivos de afecto hacia la marca, implica compromiso de recompra, pero como una buena intención podría ser una acción anticipada y no realizada.
- D) *Fidelización de acción*: Es el compromiso con la acción de recompra, a pesar de las influencias situacionales.

3.1. Tipos de Medición de la Lealtad

Según (Reyes, 2009), se distinguen entre dos formas de medición de la Lealtad de nuestros clientes: la medición paramétrica y la medición no-paramétrica.

- La Medición Paramétrica se puede aplicar en mercados o categorías de productos o servicio cuyo uso está supeditado a una renovación periódica: suscripciones, tarjetas de crédito, cuentas bancarias y seguros entre otros. Su medición es comparativamente sencilla y se basa en los registros propios de la empresa
- Por el otro lado, Reyes nos indica que en la Medición No-Paramétrica se genera un Mapa Diagnóstico de Lealtad (MADIL) mediante cuestionarios estructurados, a través de los cuales recabamos “la voz del cliente”.

3.2. Tipos de Clientes

La escalera de la fidelidad según (Martínez Rives, 1999)



Figura 2. Escalera de la fidelidad (Martínez Rives 1999)

- *Mercado potencial*: conjunto de personas que adquieren el servicio de la empresa, pero por una razón u otra no lo están haciendo, ya sea por desconocimiento de la empresa o por existencia de barreras psicológicas.
- *Prospectos*: personas que pertenecen al mercado potencial, y la empresa a identificado (nombre, dirección, etc). con el fin de iniciar una acción de venta y captación.
- *Clientes ocasionales*: personas que adquieren el servicio, como consecuencia de los esfuerzos de marketing de la empresa, deciden no seguir adquiriendo el servicio por que prefieren las ofertas de otros proveedores.
- *Clientes compartidos*: clientes que una vez hecho el primer servicio y aceptado, adquieren el servicio de forma irregular.
- *Clientes satisfechos*: son clientes que adquieren todas sus necesidades de servicio, en el mismo proveedor, debido a que esto satisface sus expectativas.
- *Promotores*: son clientes fijos de la empresa, por el alto nivel de satisfacción que recibe el servicio, se convierten en promotores del servicio, por medio de la

- comunicación boca-oído.
- *Abogados*: son promotores que llegan a convertirse en abogados, defensores de la empresa y sus productos o servicios.

(Thompson, 2002) En su libro Marketing nos muestra los 4 tipos de clientes más importantes, de los cuales se puede describir a una persona:

- **Los Apóstoles**. Son los incondicionales de nuestra empresa. Manifiestan muy elevados niveles de Encanto y de Compromiso Futuro. Son nuestros “embajadores de buena voluntad” y, por lo tanto, nuestros mejores vendedores porque dentro de sus respectivos círculos de referencia suelen ejercer un pronunciado liderazgo de opinión a favor de nuestros productos.
- **Los Clientes Terroristas**. Se caracterizan por sus muy bajos niveles de Encanto y de Compromiso Futuro. Frecuentemente han tenido una o varias malas experiencias con nuestros productos y una gran parte de ellos difunden su “mala voz” con un comparativamente elevado nivel de efectividad.
- **Los Clientes Rehenes**. Son aquellos que, a pesar de no estar contentos con nuestro producto, permanecen con nosotros. Esta situación se puede dar, por ejemplo, en mercados (semi) monopolísticos en donde no perciben alternativas viables, o en categorías donde estos clientes ven como excesivamente oneroso el costo de migración a otras alternativas.
- **Los Mercenarios**. Se caracterizan por niveles de “encanto” comparativamente elevados, pero con muy bajas intenciones de compromiso hacia nuestro producto.

3.3. Criterios de Evaluación

Se brinda un porcentaje de evaluación de 0 hasta 100 % dependiendo de cada nivel de lealtad. Esta tabla se obtuvo del libro Marketing de Thompson

Tabla 1. Criterios de Evaluación

	0-25%	25-50%	51-75%	76-100%
Lealtad				
Apóstoles				**
Mercenarios			**	
Rehenes		**		
Terroristas	**			

3.4. Modelo de Fidelización

Según Gremler y Brown (1998) este modelo es adaptado especialmente a las empresas de servicio.

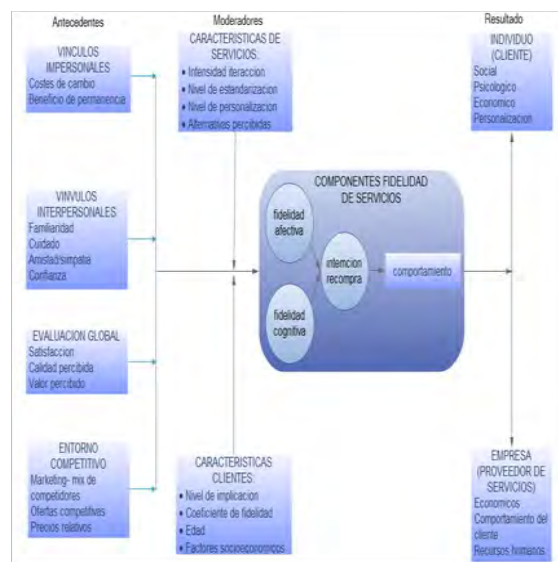


Figura 3. Modelo de Fidelización

4. Árboles de Clasificación

(Krajewski, 2000) El método del árbol de decisiones es una aproximación general a una amplia gama de decisiones de la Administración de Operaciones, como las de planificación de productos, administración de procesos, capacidad y localización.

Un árbol de decisiones es, para quien va a tomar la decisión, un modelo esquemático de las

alternativas disponibles y de las posibles consecuencias de cada una. Su nombre proviene de la forma que adopta el modelo, parecida a la de un árbol.

(Zuleyka, 2007) Un árbol de decisión puede utilizarse para la exploración de datos con uno o varios de los siguientes fines.

- Descripción: reducir una gran cantidad de datos transformándolos en una forma más compacta que preserva las características esenciales y proporciona un resumen preciso.
- Clasificación: descubrir si los datos contienen clases de objetos bien diferenciadas que puedan ser interpretadas de manera significativa en el contexto de una teoría sustantiva.
- Generalización: descubrir una relación entre variables independientes y dependientes que sea útil para predecir el valor de la variable dependiente en el futuro.

4.1. Las ventajas de un árbol de Clasificación

(Murthy, 1998) En cuanto a la tarea de clasificación, varias ventajas han sido señaladas, entre ellas:

- La adquisición de conocimiento a partir de ejemplos preclasificados salva el obstáculo que representa el adquirir el conocimiento mediante la intervención de un experto humano en el dominio.
- Los métodos basados en árboles son exploratorios en lugar de inferenciales. También son no paramétricos. Al no realizarse hipótesis sobre el modelo y la distribución de los datos, los árboles pueden modelar un amplio rango de distribuciones de datos.
- La descomposición jerárquica supone un mejor uso de las características disponibles y una mayor eficiencia computacional en la clasificación.
- Al contrario de algunos métodos estadísticos, los árboles de clasificación pueden trabajar de la misma manera con datos unimodales y multimodales.
- Los árboles llevan a cabo la clasificación a

través de una secuencia de preguntas simples, fáciles de entender, cuya semántica es intuitivamente clara para los expertos en el dominio. El formalismo de los árboles de decisión es en sí mismo muy intuitivo y atractivo para el entendimiento.

Una de las grandes ventajas de los árboles de decisión es que las opciones posibles a partir de una determinada condición son excluyentes, lo que permite analizar una situación y, siguiendo el árbol de decisión apropiadamente, llegar a una sola acción o decisión a tomar.

4.2. Por que usar Árboles de Clasificación

(Zuleyka, 2007) El trabajo sobre la inducción de árboles de decisión en estadística comienza debido a la necesidad de explorar datos de encuestas.

El Cuestionario de Lealtad, que se trabajó con la muestra de alumnos, podrá ser usado perfectamente para la creación de nuestro modelo. Se construirá una herramienta sencilla de leer y comprender.

5. Por qué usar Clementine.

(Hernandez, 2002) Clementine hace sencillo descubrir lo que ocultan sus datos. Su interfaz gráfica y sencilla, basado en procesos y flujos de datos, incrementa la productividad de los analistas, teniendo herramientas como correlación, reglas de asociación, patrones secuenciales, segmentación y clasificación; es la solución líder en minería de datos, además de la generación de informes, que le ayuda a las organizaciones a comprender el comportamiento de las personas y a predecir qué es lo que harán.

5.1. El algoritmo C5.0:

- ID3 [Quinlan86] y C4.5 son algoritmos de inducción que extraen una estructura de clasificación a partir de los datos.
- C4.5 es el sucesor de ID3, dado a que la idea básica de usar “ventanas” o subconjuntos de datos es común a ambas técnicas.

En el año 2000, Quinlan publicó C5.0, el sucesor

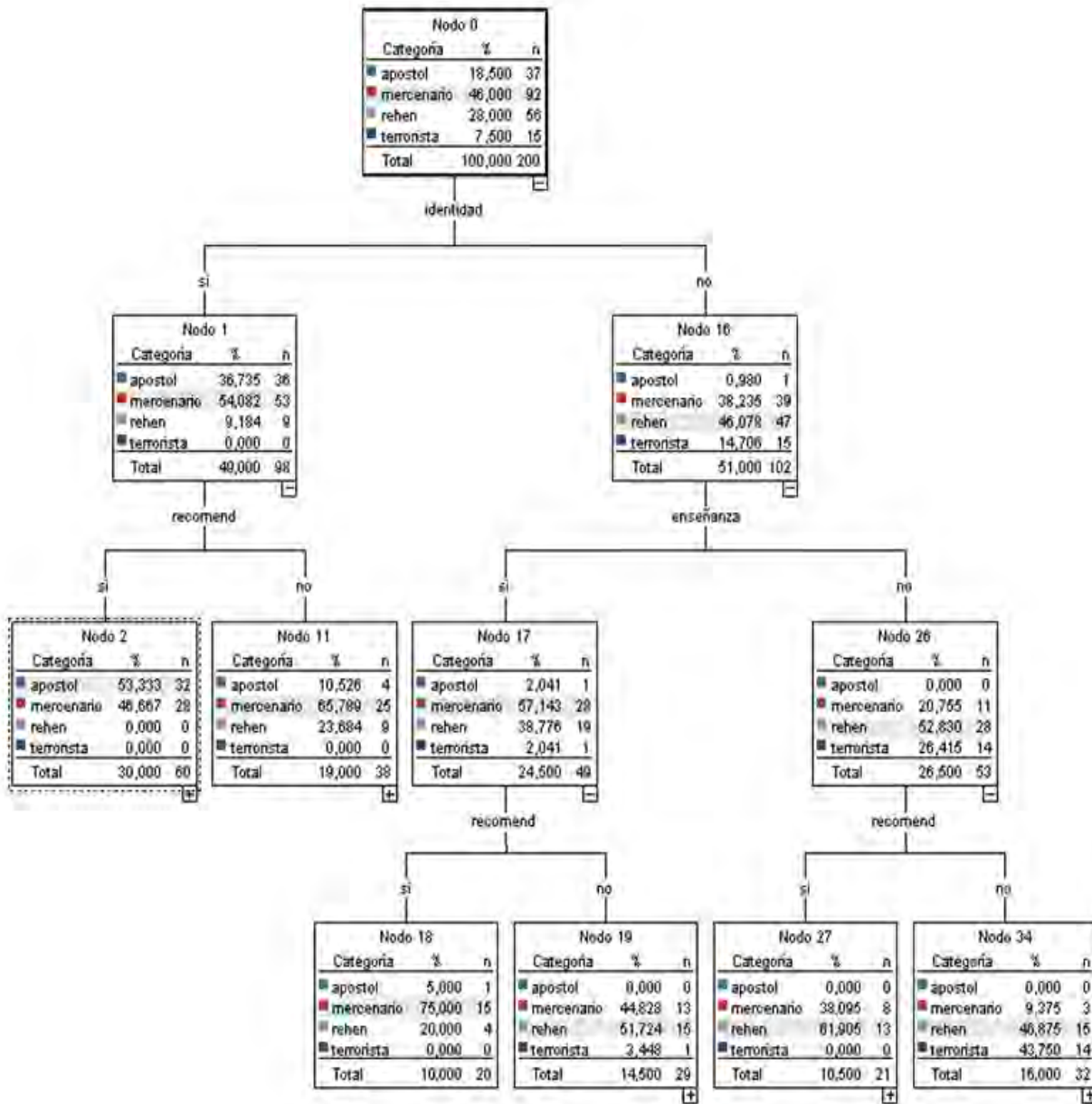


Figura 5. Árbol de Clasificación en SPSS y Clementine respectivamente

En la figura 5 se observa que la variable de entrada el 46% de los alumnos pertenece a la categoría mercenario y esto indica que la mayoría de los alumnos actualmente manifiestan niveles de encanto comparativamente elevados, pero con muy bajas intenciones de compromiso hacia la facultad, sin embargo no tienen un compromiso en el futuro para con ella.

En el AC se observa que si un alumno se siente identificado con la FIA, el alumno será un “mercenario”. Si un alumno se siente identificado con la FIA y recomienda a otros estudiar en esta facultad, el alumno pertenece a la categoría “apóstol”. Y así continuaríamos descendiendo por el árbol, hasta completar la totalidad de sus hojas (recordemos que cada hoja del árbol se refiere a la decisión a tomar).

Y así continuaríamos descendiendo por el árbol, hasta completar la totalidad de sus hojas (recordemos que cada hoja del árbol se refiere a la decisión a tomar).

Para determinar la relevancia de cada atributo con respecto a la lealtad se ha analizado el modelo AC. Para cambiar la categoría al cual pertenece un alumno se analiza los atributos ubicados en los diferentes niveles del modelo AC. Entonces, para calcular la relevancia de un atributo (Tabla N° 2) se utiliza la frecuencia de distribución, importancia relativa (IR) y la normalización (IR2, ver fórmula en la parte inferior). De acuerdo con este criterio, los atributos relevantes son: Identidad, Recomendación, Tecnología, Enseñanza.

$$IR = \frac{\text{Valor de la } I_n}{\text{Muestra}}$$

$$IR2 = \frac{\text{Valor de la } I_n}{\sum_a^b I}$$

Tabla 2. Valores de importancia

Atributo	I	IR	IR2
Identidad	200	1	0,3676470
Recomend	180	0,9	0,3308823 5
Participación	15	0.075	0,0275735 3
Tecnología	38	0.19	0,0698529 4
Enseñanza	102	0.51	0,1875
Agefac	9	0.045	0,0165441 2

La administración de la facultad podrá observar que actualmente tiene una mayoría de alumnos “mercenarios” (46%), estos son los alumnos que

están encantados pero con muy bajas intenciones de compromiso, casi el 50 %, seguido de alumnos denominados “rehenes” con un 28 % estos son los alumnos que aunque no están contentos con la facultad por diferentes razones siguen en la facultad, y un 18,5% de alumnos denominados “apóstoles” quienes están realmente comprometidos con la facultad en sus diferentes actividades y están contentos con la administración y hay un 7,5% de alumnos denominados terroristas quienes tienen muy bajos niveles de compromiso con la facultad.

Dado la relevancia de los atributos se puede observar que en el árbol generado los alumnos que tienen de 2 a menos años en la facultad son quienes más participan en las actividades , teniendo ellos mucha expectativa en su escuela académica no siendo así los que están más de dos años en la facultad. También se puede observar que los alumnos que se sienten más identificados son aquellos que se sienten satisfechos con la enseñanza que se les otorga, esto ayudará a que los alumnos recomienden a la facultad.

7.3. Identificación del Perfil de Lealtad

Tomando como base el modelo de árboles validado, se elaboró una matriz con los resultados obtenidos. Describiendo las características que presenta un alumno que clasifica para otro tipo de cliente

Lealtad	Años de Estudio	Consideran a la FIA	Recomiendan a la FIA	Participan en Actividades	Califican la Tecnología	Se identifican con la FIA	Valoran la enseñanza	
Apostol	> 2 años		SI	Ocasionalmente		SI		
				Siempre				
Mercenario	<= 2 años		SI	Ocasionalmente		SI		
				Normalmente				
				Nunca				
			NO	Aceptable				
				Anticuada				
				Moderna				
SI		NO	SI					
Menos que La mejor	NO		NO					
Rehen			NO		Mala	SI		
								Igual a La mejor
								No lo se
			SI			Igual a		NO
						Menos que		
						No lo se		
NO	La mejor		NO					
Terrorista			NO			NO	NO	
								Igual a
								Menos que
	No lo se							

Tabla 3. Matriz de Clasificación de tipos de Cliente

8. Conclusiones:

Al aplicar el modelo de AC a los niveles de lealtad que manifiesta un alumno se ha logrado establecer el perfil del alumno leal, distribuido en 4 categorías: Apóstoles, Mercenarios, Rehenes y Terroristas; encontrándose que las características de un alumno apóstol es: cuando recomienda y se siente identificado con la facultad.

Apoyándonos en la minería de datos se ha logrado obtener una herramienta que nos muestra las características más relevantes de la lealtad de un alumno, proponiendo que se trabaje primero con las características del alumno apóstol.

Referencias Bibliográficas

Vladilo M, 2008. Lealtad a grandes rasgos. España Madrid: MacGraw Hill. 730 p.

Reyes R, 2009. Satisfacción y lealtad. 2da. Ed.

México: MacGraw Hill. 420 p.

Thompson I, 2002. Marketing. 2da. ed. España: Paraninfo. 680 p.

Patrick R 2007. Lealtad y rentabilidad. Argentina Buenos aires: La pajarilla. 467 p.

Escudero J. 2008. Clementine SPSS. España Madrid: MacGraw Hill. 100 p.

Pérez C. 2007. Minería de Datos. Técnicas y Herramientas Madrid: MacGraw Hill. 808 p.

Álvarez C. 2001. Arboles de Clasificación y Minería de data. 3ra. ed. Colombia: Lily Solano. 230 p.

Gonzales R. 2008. Clasificación de los Arboles de Decisión. España MacGraw Hill. 367p.

Quinlan J. 1993. Clementine Algoritmo C5. España MacGraw Hill 457p.

Setó Pamies, D. 2004. De la calidad del servicio a la fidelidad del cliente. Madrid. ESIC. 170p.

Chiesa de Negri, C. 2009. CRM – Las cinco pirámides de Marketing Relacional. España. Deusto. 360p.

Matinez R. y Vilanova M.2004. Gestion de la Clientela – La manera de conseguir y retener clientes rentables. Madrid. ESIC. 190p.

Vierira Braga P., Ortiz Valencia L., Ramirez Carvajal S. 2008. Introducción a la Minería de Datos. E-Papers. 250p.

Krajewski, Lee J.,Larry P, 2000. Administración de operaciones. Estrategia y analisis,5ta edición. Pearson Educación.928 p.

Jiawei H., Micheline K. 2006. Data Mining – Conceptos y Técnicas. USA. Elsevier. 770p.

Murthy, S. K. (1998). Automatic construction of decision trees from data:A multi-disciplinary survey, Data Mining and Knowledge Discovery,[http://citeseerx.ist.psu.edu/viewdoc/download?](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.26.9663&rep=rep1&type=pdf)
[doi=10.1.1.26.9663&rep=rep1&type=pdf](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.26.9663&rep=rep1&type=pdf)

David F. Nettleton, 2005.Técnicas para el análisis de datos clínicos. Ediciones Díaz de Santos. 334 p.

Zuleyka Diaz Martinez, 2007. Predicción de crisis empresariales en seguros no vida mediante árboles de decisión y reglas de clasificación. Editorial Complutense. 146p.

Larraga López Pablo. 2009, Asesoramiento financiero en la Practica. España Barcelona. Editorial Bresca, 129p.