

“Diseño e Implementación de un Sistema para la Segmentación de Clientes de una Operadora Celular”

AUTORES:

Fabián Cabrera Cuenca ¹, Sergio Jonathan León García ², Ilse Lorena Ycaza Díaz ³,
Juan Aurelio Alvarado Ortega ⁴

¹ Egresado de Licenciado en Sistemas de Información 2006 e-mail:
fcabrera@conecel.com

² Egresado de Licenciado en Sistemas de Información 2006 e-mail:
sleong@conecel.com.

³ Egresada de Licenciada en Sistemas de Información 2006 e-mail:
iycaza@ecapag.com.

⁴ Director de Tópico. Ing. Juan Aurelio Alvarado Ortega, Ingeniero en Computación ESPOL, Profesor de la Facultad de Ingeniería Eléctrica y Computación de la ESPOL.

RESUMEN

Versión español:

La Operadora Celular “VIVO”, está operando desde 1993, siendo en la actualidad la empresa líder en el mercado servicios de telecomunicaciones. La compañía comercializa sus productos y servicios mediante varios canales de distribución directos e indirectos, posicionándose en el mercado como líder en la prestación de servicios de telecomunicaciones gracias a su cobertura y número de abonados.

Uno de los principales segmentos de la Operadora Celular son sus clientes postpago, los cuales tienen a su disposición una amplia gama de planes y productos que se comercializan por los diferentes canales de ventas de la empresa. La empresa siempre se encuentra en la innovación del servicio y en mantener la fidelidad de sus clientes, además de buscar siempre nuevas oportunidades de negocio.

Con este objetivo, nace la idea de un sistema informático cuya función principal sea la de dar a conocer los diferentes segmentos de clientes, usando para ello técnicas de minería de datos implementadas en un sistema informático que tome como fuente la base de facturación postpago de la operadora, logrando con ello la consecución de nuevas estrategias que cumplan con los objetivos de la empresa.

English Version:

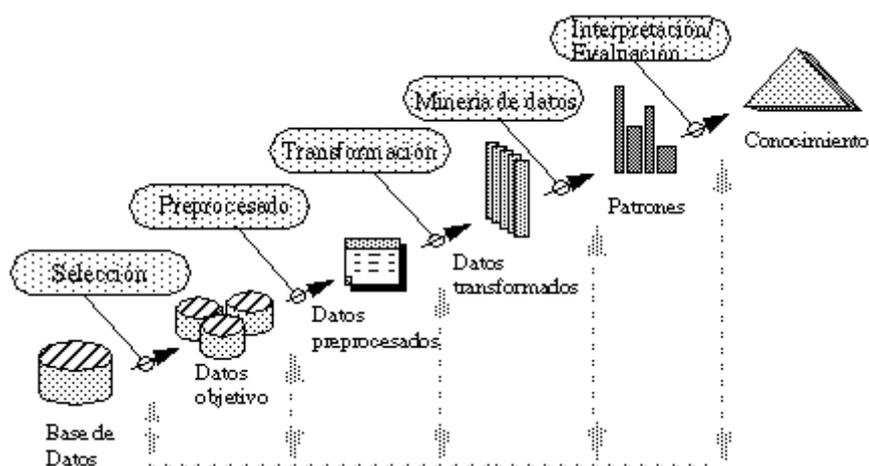
“VIVO” Cell Phone Operator, is operating since 1993, being in the actual times the lead company in the telecommunications services market. The company commercializes its products and services through several direct and indirect distribution channels, positioning in the market as the lead company in telecommunications services provisioning thanks to its coverage and subscribers number.

One of the main segments of the Cell Phone Operator is its post-paid clients, which have to their disposal a wide variety of rateplans and products that are commercialized by the many sale channels of the company.

With these objectives, borns the idea of a computer science system whose main function is to present the different segments of clients, using for that data mining techniques implemented in a computer science system that uses as source the operator’s post-paid billing data base, obtaining with that the attainment of new strategies that fulfill the objectives of the company.

INTRODUCCIÓN

La Minería de Datos o Data Mining, se encuentra dentro del conjunto de técnicas KDD (Knowledge Discovery in Databases), consiste en el conjunto de técnicas avanzadas para la extracción de información escondida en grandes bases de datos.



Las bases de datos han acumulado una gran variedad y cantidad de datos, estadísticas, índices, etc. en los cuales la información útil no es

fácil de encontrar a simple vista. Muchas empresas están interesadas en rescatar esa información y con la utilización de estas herramientas se pueden generar nuevas oportunidades de negocio. Algunas posibilidades que ofrecen estas herramientas son:

- Mejorar el funcionamiento de la organización.
- Optimizar el manejo de sus bases de datos.
- Predicción automatizada de tendencias y comportamientos.
- Obtener ventajas comerciales.
- Mejorar calidad de productos.
- Descubrimiento automatizado de modelos desconocidos.
- Descubrimiento de anomalías y acciones fraudulentas por parte de clientes.

Una de las herramientas que se utilizan con éxito en Data Mining son los Árboles de Decisión, que representa la relación existente entre la conclusión-decisión y sus atributos. Se produce un proceso de generalización de forma que el árbol de decisión generado clasifica correctamente la información de la empresa. Este árbol minimiza el número de atributos requeridos para alcanzar la conclusión-decisión y ayuda a encontrar patrones y clasificar o categorizar datos.

COMPRESION DEL PROBLEMA

Uno de los principales segmentos de la Operadora Celular son sus clientes pospago, los cuales tienen una amplia gama de planes y productos que se comercializan por los diferentes canales de ventas de la empresa.

Dado la agresiva competencia que se vive en el mercado ecuatoriano, Operadora Celular siempre se encuentra innovando en nuevos productos y tecnología, además de ofrecer promociones y descuentos que se ajusten a las necesidades del consumidor.

Con la meta de mejorar los servicios ofrecidos a los clientes del segmento pospago, se pretende mediante técnicas de minería de datos y usando la base de facturación mensual, generar información que soporte la toma de decisiones para futuros proyectos en los que se involucre innovaciones en el producto y captaciones.

Con miras de lograr un aporte que beneficie a la estrategia de la organización se han planteado para el sistema propuesto los siguientes objetivos:

Objetivos

- Determinación del comportamiento del segmento pospago celular según sus rubros de facturación mensual.
- Análisis de los segmentos significativos de clientes y enfoque en la innovación del producto con el objetivo de la conservación y fidelidad de los mismos.
- Análisis de los segmentos aislados de clientes para su proyección de expansión.
- Análisis de los comportamientos menos afines, para la determinación de posibles nuevos mercados.

CONTENIDO

1. DESCRIPCION DE LOS DATOS

El sistema de facturación se alimenta de la información mensual de los consumos de los clientes. Esta información es recopilada y agrupada dando origen a la factura del cliente.

Para la obtención de los segmentos de facturación, se necesita obtener la información de la facturación individual por cliente de los archivos de facturas generados por el sistema de facturación Pospago.

2. PREPROCESAMIENTO DE LOS DATOS

En minería de datos, los datos reales pueden ser impuros y a su vez esto puede conducir a la extracción de reglas poco útiles. La preparación de datos genera "datos de calidad, los cuales permiten obtener reglas de calidad para los datos.

Como primer paso en la generación de información consistente, se procedió a generar una nueva vista de datos, se eliminaron las columnas que no tienen relevancia por su alta cardinalidad (muchas variedades de datos únicos en la columna).

2.1 Técnicas de PreProcesamiento Empleadas

Se han seleccionado 3 técnicas, las cuales ayudarán a la depuración y normalización de los datos:

- Detección y eliminación de registros con datos aberrantes. (mediante la determinación de la distancia Mahalanobis).
- Normalización de las columnas numéricas.
- Eliminación de aquellas columnas que contienen datos únicos los cuales no aportan al objetivo del modelo.

3. SELECCIÓN Y EVALUACION DEL MODELO

Se clasificaron los datos agrupándolos por similaridad. Se apunta a una técnica de clasificación que permita generar un modelo capaz de ser implementado en las decisiones del negocio. El modelo a emplear usará un algoritmo de árbol de decisión.

Para nuestro modelo, la variable a explicar es el valor Total de los consumo del cliente, sin embargo la columna no puede ser usada en su inicial forma debido a que muestra valores continuos.

Mediante una pre-clasificación de los clientes según su rentabilidad, la cual fue obtenida en función a su consumo con respecto al promedio del plan, se logra obtener la variable CLASIFICACION la cual es explicada a continuación.

3.1 Valor Nominal para la variable determinante

El criterio de clasificación usado es la rentabilidad del cliente según su valor total de facturación por período, se pondera la clasificación al promedio obtenido por cada plan según:

Si total > promedio_plan entonces
Rentable
Caso contrario
No rentable

3.2 Reducción del error de predicción del modelo

Para disminuir el error de prueba del modelo se usa el criterio de post-poda a los nodos del árbol.

Para aplicar la post-poda de los nodos del árbol, se usa el criterio de eliminación de los nodos hijos cuyo error de prueba sea mayor al error del padre.

La determinación del error de prueba se la realiza de la siguiente manera:

$$\text{error nodo} = \frac{(f + z * z / (2 * N) + z * \text{raiz}(f / N - f * f / N + z * z / (4 * N * N)))}{(1 + z * z / N)}$$

En donde:

E = Cantidad de nodos incorrectamente clasificados

F = El porcentaje de nodos incorrectamente clasificados.

N = La cantidad de instancias de la clase

Z = el valor porcentual de la región de aceptación de la distribución normal – 69%

4. ANALISIS DE LOS SEGMENTOS

Una vez realizada la ejecución del algoritmo de generación del árbol de decisión sobre los datos correspondientes a un período de facturación, 300,000 registros, se determinaron 20 hojas de las cuales 10 representan los segmentos más importantes de los clientes y dos rubros se constituyen en los de mayor incidencia para la determinación de la rentabilidad de los clientes.

Servicios Adicionales.- Se puede notar en el modelo que los valores facturados de servicios adicionales tienen un alto impacto en la facturación total del cliente. Un servicio adicionales, es un servicio de valor agregado que presta la operadora, diferente del consumo primario de voz o datos (entiéndase por datos sms, wapy mms).

Al hacer el análisis de este rubro en específico se puede notar que indistintamente del plan los clientes siempre tienden a generar consumos que no son de voz, mensajes o datos.

Paquete de SMS.- Se puede observar además en el modelo que la segunda variable significativa para la rentabilidad de los clientes es la contratación de paquete de mensajes escritos. Este rubro aplica a todos los planes disponibles.

CONCLUSIONES

Habiendo determinado y analizado el modelo de decisión binaria obtenido para efectos de este proyecto, se concluye que los clientes pospago se encuentran claramente clasificados en dos grandes grupos:

- Clientes de categoría rentable, con alto poder de consumo, cuyo estilo de comunicación es primordialmente el servicio de voz, además de preferir consumos de servicios adicionales.
- Clientes de categoría rentable, con nivel de consumo medio, cuya preferencia de consumo se concentra en la contratación de paquetes de servicios de valor agregado (sms, mms y wap), menos facturación en llamadas y cuyo estilo de comunicación es principalmente mediante SMS.
- Basado en los datos obtenidos de la ejecución proporcionada por el sistema, aproximadamente el 15% de la clase rentable esta conformada por los clientes identificados con alto poder de consumo, y el 85% por aquellos identificados con un nivel de consumo moderado.
- La implementación de un sistema de Segmentación de los Clientes se convierte en una herramienta más de apoyo a la toma de decisiones, tanto comerciales como operacionales, que mejorarán sin duda el valor agregado del producto ofrecido.

Finalmente se recomienda, tomar acciones de mercadeo y logística, usando la información obtenida por el sistema en cada cierre de facturación; obteniendo de este modo el beneficio no solo de la expansión hacia los nuevos segmentos de clientes que se encuentran en formación, como los descubiertos durante la realización de este proyecto, si no a la vez conocer mejor a los segmentos ya definidos y afianzar la lealtad de los mismos mediante la mejora de los productos y servicios que ofrece la institución.

Ing. Juan Alvarado
Director de Tesis