

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA DE SEGMENTACIÓN DE LOS JUEGOS DE CASINOS DE ACUERDO CON SU NIVEL DE ACTIVIDAD

AUTORES

Lourdes Paucar Robles¹, Marco Paucar Robles², Roberto Alcivar García³, Juan Alvarado Ortega⁴

¹ Analista de Sistemas PROTCOM-ESPOL 2005, Egresada de Licenciada en Sistemas de Información FIEC-ESPOL 2006.

² Analista de Sistemas PROTCOM-ESPOL 2001, Egresado de Licenciado en Sistemas de Información FIEC-ESPOL 2006.

³ Analista de Sistemas PROTCOM-ESPOL 2003, Egresado de Licenciado en Sistemas de Información FIEC-ESPOL 2006.

⁴ Director de Tópico. Magíster en Administración Pública con Mención en Liderazgo ESPOL, Ingeniero en Electricidad Especialización Computación ESPOL, Profesor de ESPOL desde 2002.

RESUMEN

El sector al cual apunta nuestro proyecto es el sector de la industria del Juego, específicamente en el sistema de las máquinas tragamonedas, las cuales bajo la perspectiva de un modelo de minería de datos permita a los administradores obtener conclusiones a fin de aplicarlas en su modelo de negocios.

Específicamente tratamos el negocio de la empresa Interecuador S.A. dedicada al negocio de explotación de máquinas tragamonedas en calle, en locales y en salas de juego y casinos. El negocio posee diversos productos que pueden ser catalogados bajo dos grandes categorías: Mesas y máquinas.

Dentro de mesas abarca todos los juegos de azar basados en cartas que tienen la intervención humana, este tipo de productos es exclusivo de casinos.

Dentro de máquinas existen múltiples variedades de marcas, modelos y juegos los cuales se utilizan tanto en el negocio de calle, locales, casinos y salas de juego.

Dadas las diversas características de los mismos es necesaria la implementación de un método que nos permita explorar los datos y encontrar la mejor combinación de características de determinada máquina de manera que su implementación en determinado local basado en la estadística me garantice la mejor producción de la misma.

INTRODUCCIÓN

Minería de datos es la extracción de información oculta y predecible de grandes bases de datos, es una poderosa tecnología con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información transaccional o Data Warehouse. Las herramientas de la minería de datos predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones pro activa y conducida por un conocimiento acabado de la información. Los análisis prospectivos automatizados ofrecidos por un producto así van más allá de los eventos pasados provistos por herramientas retrospectivas típicas de sistemas de soporte de decisión. Las herramientas de Minería de Datos pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información casi no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

CONTENIDO

1. DESCRIPCIÓN DE LA INSTITUCIÓN

1.1 DESCRIPCIÓN DEL NEGOCIO

Nuestro modelo de negocio esta enfocado a empresas dedicada al manejo de salas de juego y casinos, con múltiples salas y variadas combinaciones de maquinas, modelos y juegos que le proporcionen al negocio la mayor rentabilidad.

1.2 MODELO DE NEGOCIO

El modelo de negocio se basará principalmente en la toma de decisiones de la mejor combinación de máquinas para obtener la mayor rentabilidad posible de una determinada sala de juego, a la vez de encontrar otros nichos de mercado donde permita la expansión del negocio, y seleccionar determinadas características que justifican la rentabilidad de dicha máquina basado en un criterio de selección de las mismas.

1.3 CADENA DE VALOR DE NEGOCIO

La cadena de valor del negocio en el cual se basa el siguiente proyecto es la siguiente:

Compra – Configuración – Instalación – Ventas - Revisión

En cada uno de estos pasos se llevan importantes decisiones que abarcan todo el modelo de negocio:

- **Compra:** Proceso mediante el cual se lleva a la licitación de la compra de maquinas de juego, ya sea mediante la concertación de la forma de pago o en su caso el tipo de participación. Ya sea compartida o por porcentaje.
 - **Compartida:** Las maquinas pertenecen a una empresa que recibe un valor por tener las mismas en un local de otra empresa
 - **Porcentaje:** La empresa dueña de las maquinas recibe un porcentaje del valor recaudado por juego en dichas maquinas
- **Configuración:** Una vez gestionada la compra de máquinas se procede a la configuración de las mismas para que procedan a funcionar bajo el esquema indicado en la maquina. Esta parte es neurálgica debido a que es muy importante la determinación de la fórmula de juego de la máquina dado que el futuro cálculo de la producción de la misma es determinada por dicha fórmula.
- **Instalación:** Una vez determinada la fórmula de juego de la máquina y su cálculo de producción es la instalación de la máquina un proceso hasta este momento totalmente empírico en el cual basado en la experiencia del departamento de operaciones se determina el destino de dicha máquina. Es aquí que es necesaria la implementación de una solución basada en minería de datos que ayude a determinar la mejor opción para determinado local con determinada máquina, ese es el objetivo primario de este proyecto.
- **Ventas:** Este proceso es la producción de la máquina, basado en varios factores tales como la ubicación de local, nivel de juego, tipo de maquina, etc. Se procede a hacer una hoja de ruta en la cual se determina que máquinas serán recaudadas en este proceso se realiza la toma de contadores y la toma física de valores producidos por la máquina, procediendo al cuadro de la misma y su puesta en juego nuevamente. En este punto es cuando bajo un esquema tradicional el departamento de operaciones puede evaluar la efectividad de la toma de decisiones tanto como la elección, configuración e instalación de la máquina, bajo un esquema de Data Mining este paso seria una simple verificación de los resultados arrojados por la minería de datos comprobando la misma a través de su producción.
- **Revisión:** bajo el punto de revisión en la cadena de valor se pueden tomar en cuenta varios puntos significativos antes de determinar una máquina por revisión

- Revisión por fallas técnicas: máquinas con comprobadas fallas técnicas
- Retiro de máquina por cancelación de contrato
- Reconfiguración de máquina

Una vez determinada la cadena de valor es importante destacar que el proceso de selección de máquinas, su configuración e instalación al igual que su revisión son basados en un método empírico en el cual basado en el conocimiento adquirido por el departamento de operaciones se toman decisiones que afectan directamente al modelo de negocio.

2. DESCRIPCIÓN DE LOS DATOS

Una vez determinado cuales son las reglas de negocio y la operación del mismo, empezaremos a experimentar ciertos procesos que afectarán directamente al modelo de datos.

Estos procesos afectan la parte de preprocesamiento de datos, pero es necesaria antes de este paso la descripción de los datos mediante la cual se ajusta el modelo a la realidad que es.

2.1 MODELO ENTIDAD RELACIÓN

La estructura de la base de datos utilizada por el Sistema de Segmentación de los Juegos de Casino es la que se utilizo para la implementación del sistema.

2.2 SELECCIÓN DE DATOS

En un proyecto de Data Mining esta fase incluye la adquisición de diversos datos que pueden residir en una o mas base de datos, pudiendo ser una base de datos transaccional, un Data Warehouse o un Data Mart, para este proyecto se trata de una base de datos transaccional.

3. PREPROCESAMIENTO DE LOS DATOS

3.1 PREPARACION DE LOS DATOS

Una vez que ya se han seleccionado los datos, es necesario realizar algunas transformaciones. Esta etapa que corresponde a la consolidación de los datos en una sola base de datos, considera la identificación y corrección de inconsistencias en la definición de los datos, diferencias de codificación y resuelve la discordancia de múltiples valores para un mismo dato.

Se requiere "limpiar" los datos, uniendo o mezclando las fuentes de los datos y la posible aparición de nuevas columnas o campos. El resultado final es normalmente dependiendo de la herramienta de minería utilizada lista para la aplicación del Data Mining.

Pasos a seguir para la preparación de los datos:

- Análisis de todas las variables y de su tipo de dato
- Creación de nuevas variables o reducción de las mismas
- Selección de las variables dependientes
- Poblar los campos vacíos (NULL) con la clase mas probable
- Detección de los datos aberrantes
- Reducción considerable de las columnas utilizando Software para Data Mining
- Numerización o Standardización de las columnas numéricas y que lo requieran
- Disminución del error en la selección de variables

3.2 DICCIONARIO DE DATOS DEL ARCHIVO DE MINERÍA

Realizado el preprocesamiento de los datos la solución nos da una estructura de datos la cual determinan las principales entidades de mi modelo de datos.

4. SELECCIÓN Y EVALUACIÓN DEL MODELO

4.1 MODELADO EN MINERÍA DE DATOS

El nombre de Data Mining deriva de las similitudes entre buscar valiosa información de negocios en grandes bases de datos por ej.: encontrar información de la venta de un producto entre grandes montos de Gigabytes almacenados y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una inmensa cantidad de material, o investigar inteligentemente hasta encontrar de forma exacta donde residen los valores. Dadas bases de datos de suficiente tamaño y calidad, la tecnología de Data Mining puede generar nuevas oportunidades de negocios al proveer estas capacidades:

Predicción automatizada de tendencias y comportamientos. Data Mining automatiza el proceso de encontrar información predecible en grandes bases de datos.

Las bases de datos pueden ser grandes tanto en profundidad como en ancho: más columnas. Los analistas muchas veces deben limitar el número de variables a examinar cuando realizan análisis manuales debido a limitaciones de tiempo. Sin embargo, variables que son descartadas porque parecen sin importancia pueden proveer información acerca de modelos desconocidos. Un Data Mining de alto rendimiento permite a los usuarios explorar toda la base de datos, sin preseleccionar un subconjunto de variables.

Más filas. Muestras mayores producen menos errores de estimación y desvíos, y permite a los usuarios hacer inferencias acerca de pequeños pero importantes segmentos de población.

4.2 CONSTRUCCIÓN DEL MODELO

Un modelo es una abstracción de la realidad. No puede predecir todos los eventos individuales, pero un buen modelo es una guía útil, para entender la organización y sugerir acciones que permitan lograr el éxito.

4.3 ÁRBOLES DE DECISIÓN COMO ALGORITMO DEL MODELO

El algoritmo de árboles de decisión aplicado en el modelo de este proyecto de minería de datos, ya que las estructuras de forma de árbol representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos enfocado en conocer que factores influían directa e indirectamente en obtener mayores y/o menores rentabilidad en la salas de casino.

Un modelo implementado utilizando árboles de decisión nos permite establecer cual es el nivel de influencia de cada una de las características consideradas para este modelo en obtener la mayor rentabilidad en la sala de casino.

4.4 EVALUACIÓN DEL MODELO SELECCIONADO

El modelo de minería de datos se aplicó a 8486 registros, después de utilizar las técnicas del preprocesamiento quedaron 8390 registros puros, es decir la data inconsistente o datos aberrantes fueron 96. La información proviene de la recolección diaria de los contadores de cada maquina en las diferentes salas de juegos que corresponden a los meses de Marzo, Abril, Mayo.

4.5 DESPLIEGUE DE PATRONES

Posterior a la construcción del modelo y al descubrimiento de los patrones, se debe evaluar sus resultados e interpretar su significado. Cuando se realiza la validación de los patrones, se encuentra una razón de exactitud. Es importante recordar que esta razón se aplica solo a los datos con los cuales se construyó el modelo. En la práctica la exactitud puede variar si la muestra de datos a las que se les aplico el algoritmo de descubrimiento no representa significativamente la globalidad de los datos.

La segmentación de los datos se la presentó de la siguiente manera y esta dado de acuerdo a la rentabilidad es decir ALTA RENTABILIDAD y BAJA RENTABILIDAD.

5. IMPLEMENTACIÓN DEL SISTEMA

5.1 SEGMENTACIÓN DE JUEGOS DE CASINO

Una de las opciones del sistema brinda la posibilidad de en cualquier momento ejecutar el proceso de generación del archivo de minería de datos de acuerdo a ciertos parámetros, lo que nos permitirá actualizar la información que sirve de entrada para la ejecución del modelo de minería lo cual también esta disponible. De esta manera se puede realizar un análisis de los datos donde periódicamente podemos analizar datos nuevos o nos brinda la posibilidad de aplicar ciertos criterios de selección de datos que permitirá aplicar un sesgo a los datos para algún trabajo particularmente focalizado.

La periodicidad con que se realice este proceso de actualización del archivo de minería de datos y su posterior evaluación por el modelo, debe ser definida.

5.2 MONITOREO DEL MODELO

Tal como cambian los procesos de negocio durante el tiempo, la validez de los patrones descubiertos a partir de los datos históricos pueden verse deteriorados. Es muy importante detectar estos cambios lo mas temprano posible, a través del monitoreo constante con relación a los nuevos datos. Cambios significativos a los patrones apuntará a la necesidad de descubrir nuevos patrones a partir de los nuevos datos.

5.3 BENEFICIOS Y PROBLEMAS DEL PROCESO DE DATA MINING

La aplicación del proceso de minería de datos puede significar para las organizaciones obtener una serie de beneficios, así como también su ejecución se puede ver dificultada por ciertos problemas que deben ser resueltos previamente, ya que de manera contraria no se podrían obtener buenos resultados.

5.3.1 BENEFICIOS

Extraer información táctica y estratégica almacenada en sus Data Warehouse (o Data Mart). O bien, puede orientar la creación de una base de datos Ad hoc.

La búsqueda de patrones la realiza en forma automatizada, por medio de herramientas que ejecutan el descubrimiento, generando modelos mediante la aplicación de diferentes técnicas, tales como árboles de decisión.

Existe gran diversidad de herramientas de Data Mining en el mercado, lo que facilita la aplicación de este proceso y aumenta el nivel de confianza de sus resultados, debido a la aplicación automatizada.

No requiere de personal experto en ciencias estadísticas o altamente entrenadas en estas ciencias, por el contrario, es suficiente un especialista en

análisis de datos y un experto en las áreas de la organización, que tenga conocimientos básicos en áreas estadísticas que le permita entender y aplicar los resultados.

5.3.2 PROBLEMAS

La falta, por parte de las organizaciones de un repositorio histórico, especialmente un Data Warehouse o Data Mart, además de no contar con base de datos bien definidas, con datos íntegros, consistentes, es decir, con datos en condiciones para la aplicación del Data Mining.

En otras palabras, la falta de una cultura informática y visión corporativa, juegan un papel negativo al momento de aplicar el proceso de data Mining. En organizaciones con estas características, este proceso puede requerir más tiempo, pero con una adecuada reestructuración de las bases de datos, creación de buenos repositorios y una serie de medidas, se puede obtener buenos resultados pero con un gasto mayor de tiempo.

CONCLUSIONES

Dado el requerimiento de obtener datos y patrones basados en un sistema que genera información muy valiosa para el negocio es determinar cuales son las mejores combinaciones de características tanto de juego como de máquina para maximizar el ingreso basado en sus diferentes ubicaciones geográficas.

Pero mucho más importante es determinar o predecir patrones de comportamiento que no permitan basarse en el pasado para tomar decisiones sino en base a hechos determinar el futuro.

A continuación explicamos uno de los patrones que indican que la siguiente combinación de características la hacen apta para obtener mayor rentabilidad.

ID	NODO	CLASE	CONDICION
1	34	bueno	$botmaq > 7$ $matrizjuego = 7$ $diasemana \leq 2$ $ubic > 2$

La condición del nodo nos indica que para obtener mejores resultados debemos de seleccionar la siguiente combinación:

- Botones por máquina debe ser mayor a 7 ($botmaq > 7$).
- La matriz de juego (matrizJuego) debe ser igual a 7=(2x5)
- Los días que predicen mayor juego debe ser menor igual a 2 (Lunes y Martes).
- La ubicación ideal debe ser mayor a 2: (Centro y Sur).

REFERENCIAS

1. Wiley, DataMinig Techniques (2da. Edición, Michael J.A. Berry and Gordon S. Linoff)
2. <http://www.cs.waikato.ac.nz/ml/weka>
3. http://www.cs.waikato.ac.nz/ml/weka/index_downloading.html

Ing. Juan Alvarado O.
Director de Tesis