

Introducción: Data Mining

¿ Qué es la minería de datos ?

Extracción de **modelos y patrones** interesantes, potencialmente útiles y no triviales desde bases de datos de gran tamaño.

Conceptos de Data Mining

Modelo: Representación abstracta de un conjunto de datos.

Patrón: Similar a un modelo pero se enfoca en un subconjunto de los datos.

Un algoritmo en minería de datos (o aprendizaje automático) es un conjunto de heurísticas y cálculos que permiten crear un modelo a partir de datos

El modelo de minería de datos que crea un algoritmo a partir de los datos puede tomar diversas formas, incluyendo:

- Un conjunto de clústeres que describe cómo se relacionan los casos de un conjunto de datos.
- Un árbol de decisión que predice un resultado y que describe cómo afectan a este los distintos criterios.
- Un modelo matemático que predice las ventas.
- Un conjunto de reglas que describen cómo se agrupan los productos en una transacción, y las probabilidades de que dichos productos se adquieran juntos.

¿ Para qué sirve la minería de datos?

La minería de datos puede ayudar a las organizaciones a tener una visión útil de su negocio a partir de los datos que se ha recopilado en los últimos años y tomar mejores decisiones para alcanzar nuevas metas.

Técnicas de minería de datos

SQL Server en la minería de datos incluye los siguientes tipos de algoritmos:

- **Algoritmos de clasificación** , que predicen una o más variables discretas, basándose en los demás atributos del conjunto de datos.
- **Algoritmos de regresión** , que predicen una o más variables numéricas continuas, como pérdidas o ganancias, basándose en otros atributos del conjunto de datos.
- **Algoritmos de segmentación** , que dividen los datos en grupos, o clústeres, de elementos que tienen propiedades similares.
- **Algoritmos de asociación** , que buscan correlaciones entre diferentes atributos de un conjunto de datos. La aplicación más común de esta clase de algoritmo es la creación de reglas de asociación, que pueden usarse en un análisis de la cesta de compra.
- Los **algoritmos de análisis de secuencias** resumen las secuencias frecuentes o episodios en los datos, como una serie de clics en un sitio web o una serie de eventos de registro que preceden al mantenimiento del equipo.

Minería de Datos

El Data Mining (Minería de datos), es el conjunto de técnicas que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de dichos datos en un contexto.



Etapas principales



01
DEFINIR EL PROBLEMA

Trazar los objetivos que el cliente desea.



02
PREPARAR LOS DATOS

Selección, enriquecimiento y transformación de las bases de datos.



03
DETERMINAR EL MODELO

Según los objetivos planeados se genera el modelo o los modelos usando los conocimientos adquiridos en la etapa anterior.



04
ANÁLISIS DE LOS RESULTADOS

Verifica si los resultados obtenidos son coherentes y novedosos y si aportan un nuevo conocimiento que permita considerar decisiones.

Aplicaciones

PRONOSTICO

Pronóstico de demanda, ventas, producción y scoring crediticio.

RIESGO Y PROBABILIDAD

Modelos estadístico, descriptivo, probabilísticos y simulaciones.

RECOMENDACIONES

Segmentaciones, clusterizaciones, georeferenciación, análisis sectoriales y análisis de plazas.

AGRUPACIÓN

Distribución de clientes o eventos en grupo de elementos relacionados, análisis y predicción de afinidades.

Tipos



PREDICTIVO

La minería de datos predictiva, como su nombre indica, intenta predecir o responder preguntas futuras con base en un estudio del comportamiento pasado

DESCRIPTIVO

Este tipo de minería, trata de proporcionar información entre las relaciones de los datos y sus características.



Implementación Data Mining en SQL Server 2008 y Visual Studio

1.- Ejecutar SQL Server 2008 (Base de datos a utilizar es BIBDTOP).

2.- Ejecutar Visual Studio y seleccione Analysis Services Project:

Nombre : DataMining_NombreAlmuno

3.- Data Source (Orígenes de datos): Crear la conexión al servidor a utilizar, presionamos NEW y colocamos serverbd /localhost, finalmente seleccionamos la base de datos para realizar el ejercicio BIBDTOP, no olvide presionar Test Connection.

4.- En la ventana Data Source Wizard, presione Next, y en el apartado de Impersonation Information no olvide seleccionar "Use the service account" para poder implementar el proyecto final. En Data source name, coloque el nombre de la base de datos a utilizar BIBDTOP. Presione Finish.

5.- Data Source View (Agregar vista al origen de datos): Seleccione la base de datos y en el paso siguiente arrastre las tabla a utilizar para el ejercicio DMCompras (Compras que posiblemente realizará un usuario a futuro), presione next y finalmente finish.

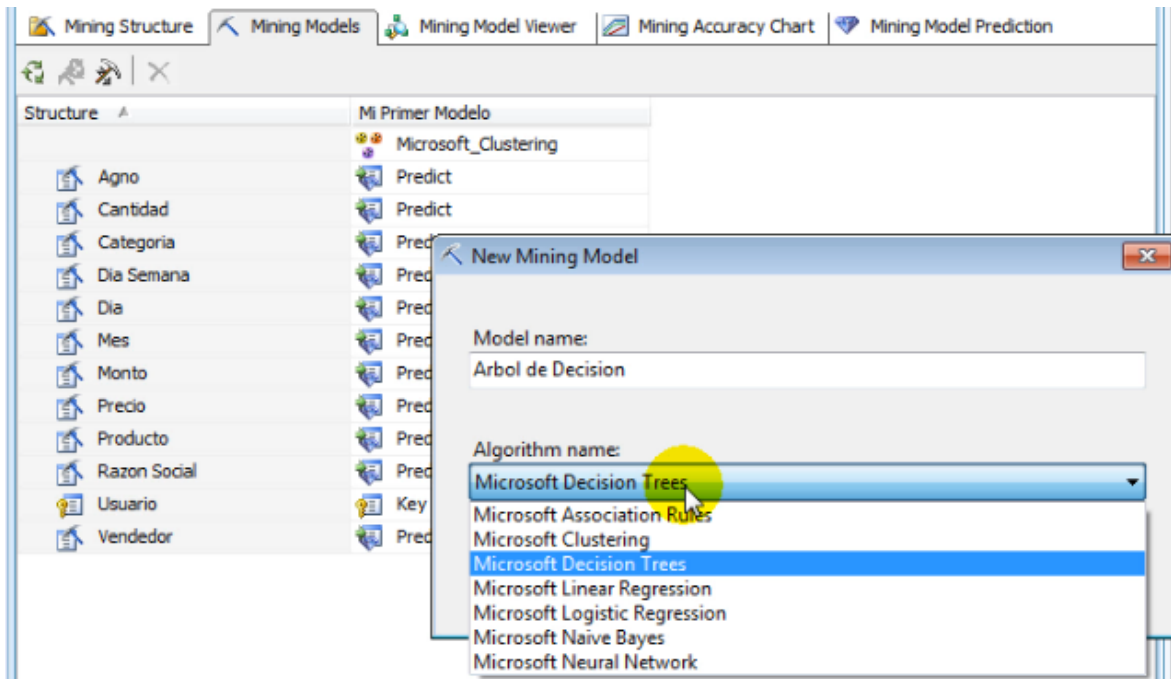
6.- En Mining Structures, realice un click derecho y seleccione New Mining Structure, presione next y seleccione en:

- Select the Definition Method: From existing relational database or data warehouse
- Create the Data Mining Structure: Microsoft Clustering (También se puede utilizar otros)
- Select Data Source view: BIBDTOP
- Specify Table TYPES: DMCompras (case check)
- Specify the Training Data: seleccione todo los campos en input y Prediction a excepción de usuario que se utilizará como key
- Specify Columns: Presione Next
- Create Testing Set / Percentage of data for testing: seleccione 80%
- Completing the Wizard: Mining structure name (Clustering), Mining model name DM Compras (Active el check Allow drill trough).

7.- En el explorador de soluciones, precisamente en DataMining realice un click derecho y seleccione propiedades, Propiedades de configuración seleccione Deployment / Target / Database : DM_NombreAlumno. Finalmente presione aceptar.

8.- Implementar el proyecto desde DataMining.

9.- En la pestaña Mining Models, debe crear un New Mining Model (seleccione el martillo).

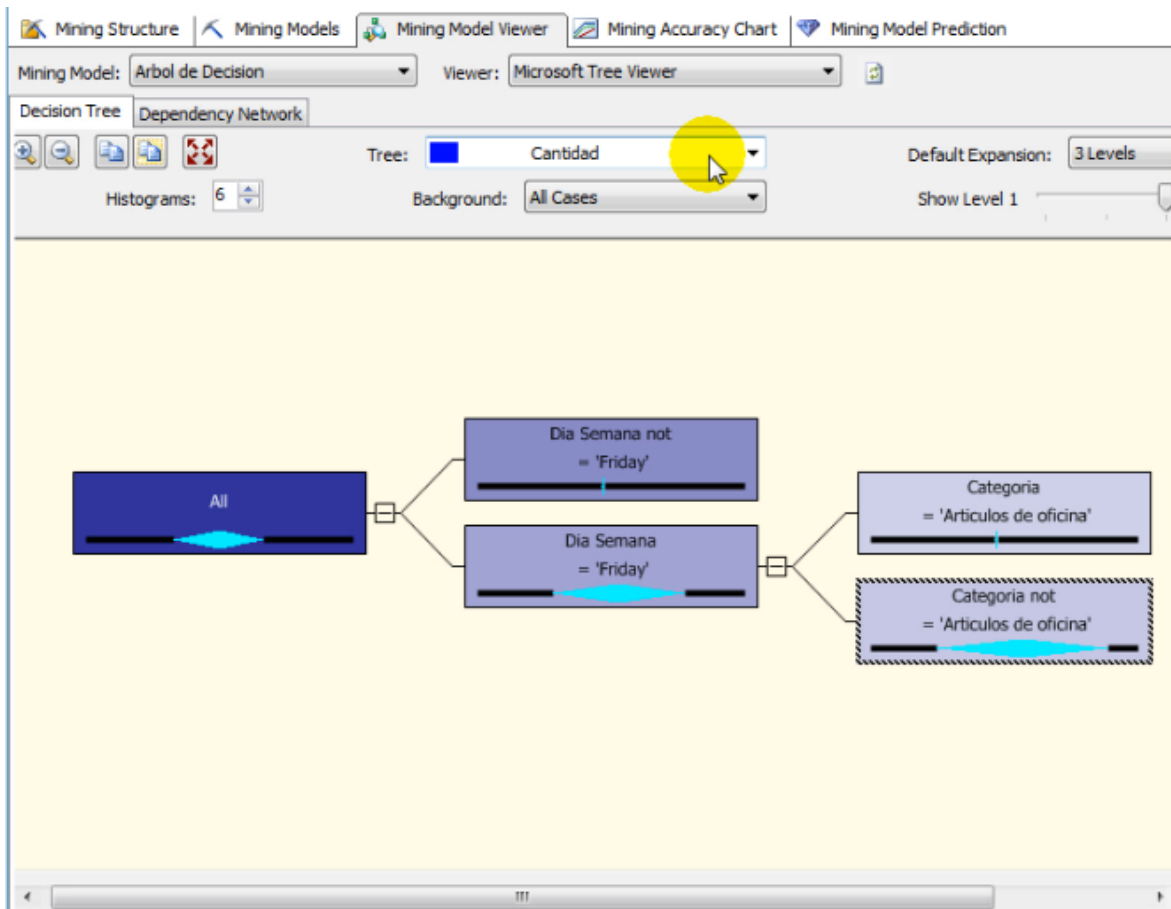


Finalmente presione las flechas de color verde para implementar el modelo de minería de datos.

10.- Mining Model Viewer.

Mining Model : Arbol de Decisión.

Viewer: Microsoft Tree Viewer.

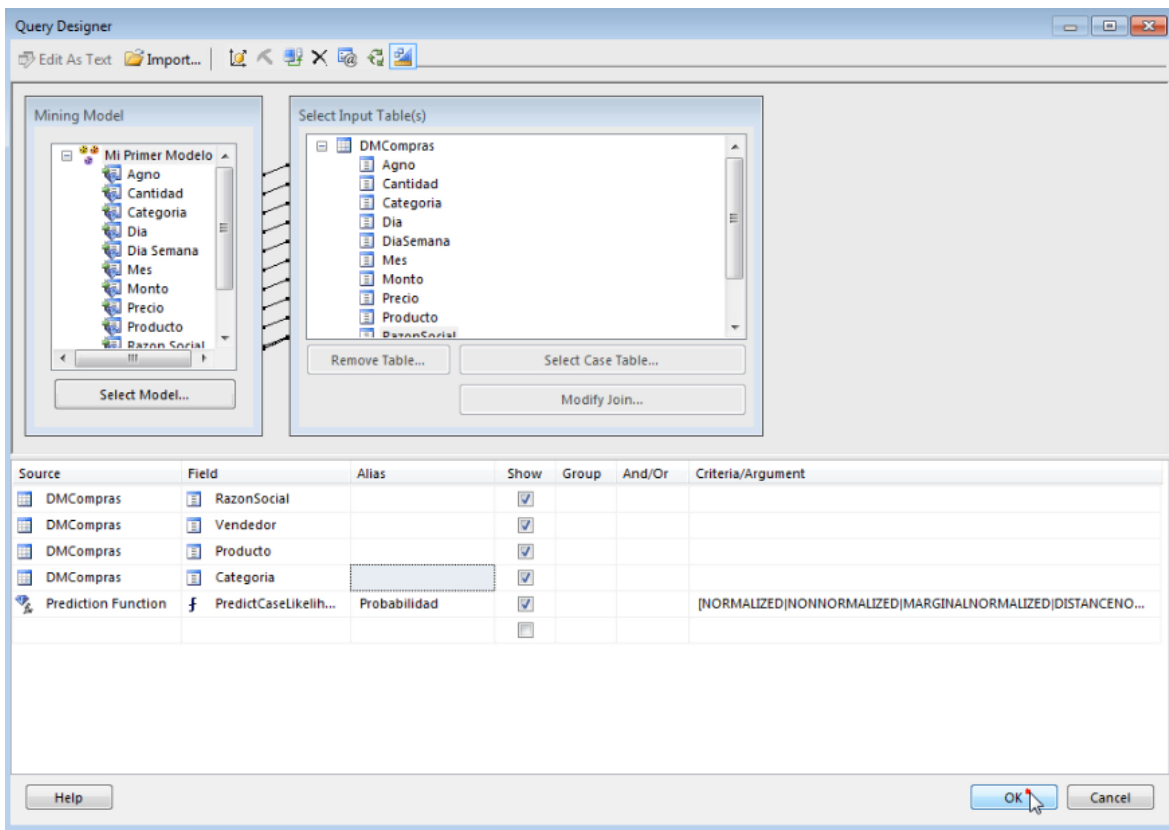


11. Visual Studio: Creamos un Report Server Project

- Shared Data Source: Add New Data Source.
- Name: DSServer.
- Type : Microsofr SQL Server Analysis Services .
- Edit (Server name): serverbd.
- Database: DM_NombreAlumno.

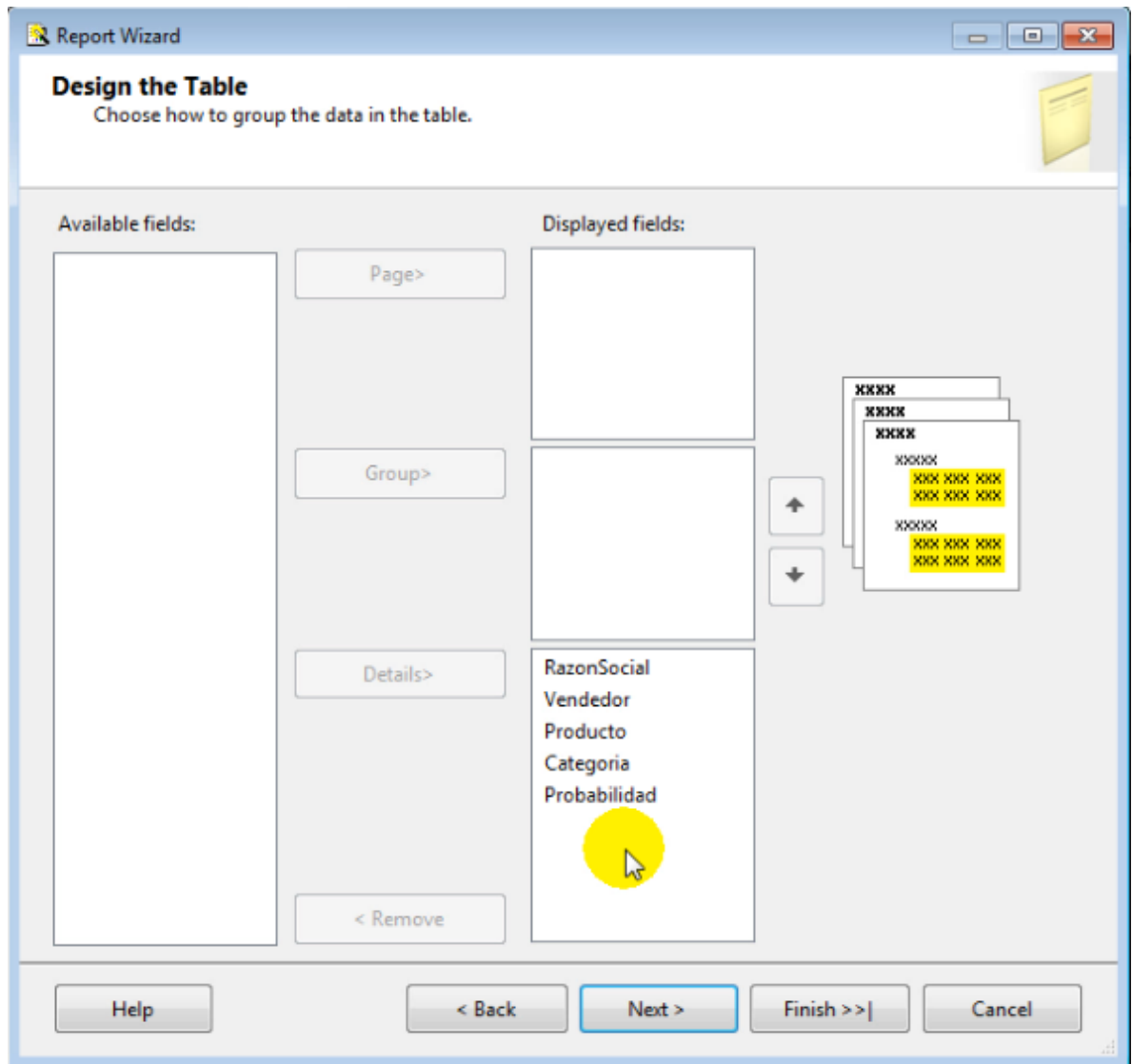
12. Report.

- Shared data source : DSServer .
- Query Builder.
- Mining Model : Seleccione Mi Primer Modelo.
- Select Input Table : DMCompras



Presione ok.

- Seleccione el reporte Tabular.



- Seleccione un nombre para su reporte y finalmente implemente la vista en Preview.