# 1.3. Herramientas Comerciales de Análisis de Datos

## **KnowledgeSeeker de Angoss Software International, Toronto, Canada**

Puntos Clave:

* Herramienta interactiva de clasificación.
* Basada en los algoritmos de árboles de decisión CHAID y XAID.
* Se ejecuta sobre plataformas Windows y UNIX

Ventajas:

* Representación flexible de árboles de decisión.
* Provee características para permitir la identificación de la relevancia de los resultados en los negocios.
* El API permite usar los resultados del análisis en aplicaciones personalizadas.

Aspectos a tener en cuenta:

* Solo soporta árboles de decisión
* Poco soporte para la transformación de datos.
* El soporte para predicción se limita a la exportación de las reglas generadas.

Cuando usarla:

* Si se necesita una herramienta que permita adelantar una visión instantánea general de sus datos.
* Si necesita una herramienta interactiva para explorar sus datos.
* No está indicada si se necesita una herramienta que soporte predicción desde dentro de sus datos.

## **DataCruncher de DataMind, San Mateo, CA, USA**

Puntos Clave:

* Herramienta de Data Mining para clasificación y clustering
* Basada en Tecnología de agentes de redes (ANT Agent Network Technology)
* La aplicación servidor se ejecuta sobre UNIX y Windows NT; la aplicación cliente en todas las plataformas Windows.

Ventajas:

* Fácil de usar, ya que los modelos necesitan pocas adaptaciones.
* Agent Network Technology puede ser utilizada para clasificación, predicción y clustering no supervisado.
* Resultados versátiles, que permiten una minuciosa valoración de los modelos y de sus resultados

Aspectos a tener en cuenta:

* Se necesita familiarizarse con la tecnología para comprender los resultados.
* Está basada en una técnica propietaria
* Tiene soporte limitado para la transformación de datos.

Cuando usarla:

* Si se necesita una herramienta cliente-servidor con una interface fácil de usar.
* Si se necesita valorar para cada caso la bondad de la predicción de los modelos.
* Si quiere invertir algún esfuerzo en hacer un completo uso del análisis de resultados.

## **Intelligent Miner de IBM, Armonk, NY, USA**

Puntos Clave:

* Soporta múltiples operaciones de data minino en un entrono cliente servidor
* Utiliza redes de neuronas, árboles de inducción y varias técnicas estadísticas.
* Trabaja sobre clientes Windows, OS/2 y X-Windows, y servidores AIX (incluyendoSP2), OS/400 y OS/390.

Ventajas:

* Buen soporte para análisis de asociaciones y clustering (incluyendo visualización de clustering), además de clasificación y predicción.
* Optimizada para data minino en grandes bases de datos(del orden de gigabytes) ya que se aprovecha de la plataforma de procesamiento paralelo PS2 de IBM.
* Tiene un entorno de trabajo integrado con características muy interesantes tanto para usuarios expertos como no especialistas.

Aspectos a tener en cuenta:

* Algunos problemas que tenía han sido resueltos con la nueva interface que ha sido desarrollada completamente en Java.
* Solo trabaja sobre plataformas IBM, y el acceso a los datos se limita a las bases de datos DB2 y a ficheros planos.
* Inicialmente la mayoría de los proyectos requerirán entradas importantes desde los servicios de soporte y consultoría de IBM

Cuando usarla:

* Debería ir a una tienda de IBM para observar la funcionalidad del data mining integrado en su entorno de soporte a las decisiones
* Para grandes proyectos de data mining, en particular cuando los datos están contenidos en DB2.
* Si se desan utilizar varias operaciones de data mining, tales como clasificación, clustering y análisis de asociaciones.
* Para realizar análisis de cesta de la compra con varios gigabytes de datos.
* Si interesa utilizar los servicios de consultoría de IBM.

## **Clamentine de Integral Solutions, Basingstoks, UK**

Puntos Clave:

* Herramienta con un entrono de trabajo que soporta todo el proceso de data mining
* Ofrece árboles de decisión, redes de neuronas, generación de reglas de asociación y características de visualización.
* Se ejecuta sobre VMS, UNIX o Windows NT.

Ventajas:

* Interface gráfica intuitiva para programación visual.
* Las técnicas de data mining pueden complementarse combinándose entre si.
* Visión interactiva de las relaciones entre las variables a través de grafos de red.

Aspectos a tener en cuenta:

* No soporta Windows nativo.
* Es necesario familiarizarse con la herramienta para conseguir una óptima utilización de sus funcionalidades.
* No está optimizada para arquitecturas en paralelo.

Cuando usarla:

* Si se necesita una herramienta que cubra por completo el rango de los procesos de data mining.
* Si se desean combinar herramientas y modelos para construir los procesos de data mining que exijan tales requisitos.
* Si se desea desarrollar el modelo en C.
* Si se necesitan grandes capacidades analíticas y de gestión de datos sin requerir un extenso análisis de datos ni experiencia en tecnologías informáticas.

## **Alice de Isoft SA, Gif sur Yvette, Francia.**

Puntos Clave:

* Herramienta de escritorio para data minino interactivo.
* Se basa en tecnología de árboles de decisión.
* Se ejecuta sobre plataformas Windows.

Ventajas:

* La representación altamente interactiva permite guiar el análisis.
* La opción de generar gráficos provee una visión general de los datos en todas las etapas del proceso de Data Mining.
* Se trata de una herramienta económica valida para usuarios que comienzan a realizar data mining.

Aspectos a tener en cuenta:

* No tiene opciones para desarrollar modelos.
* Pequeño soporte para transformación de datos.
* No genera conjuntos de reglas optimizadas desde los árboles de decisión.

Cuando usarla:

* Si se desea usar data mining para buscar patrones y relaciones en los datos.
* Si se quiere tener la posibilidad de dirigir el análisis interactivamente.
* Si no se es un experto en data mining y se desea realizar el análisis.
* Si se quiere entender los patrones que se encuentran en la base de datos y no se desea construir modelos predictivos.

## **Decisión Series, de NeoVista Software Cupertino CA, USA.**

Puntos Clave:

* Herramientas para múltiples operaciones de data mining para el desarrollo de modelos basados en servidores.
* Proporciones algoritmos de redes de neuronas, árboles y reglas de inducción, clustering y análisis de asociaciones.
* Trabaja sobre sistemas UNIX mono o multi-procesadores de HP y Sun. Accede sólo a ficheros planos, aunque posiblemente las últimas versiones ya trabajaran contra bases de datos relacionales.

Ventajas:

* Soporta un gran rango de operaciones y algoritmos de data mining, la mayoría de los cuales han sido altamente optimizados para obtener altos rendimientos.
* Está optimizado para plataformas que trabajan en paralelo con grandes conjuntos de datos.
* Ofrece una considerable flexibilidad para construir modelos de alto rendimiento para aplicaciones de usuario final embebidas.

Aspectos a tener en cuenta:

* Las herramientas de desarrollo gráfico son bastante básicas.
* Poco soporte para la exploración de datos.
* La mayoría de los clientes necesitaran un considerable soporte de consultas para generar aplicaciones y ejecutarlas. Es necesario tener conocimientos de análisis de datos y de utilización de UNIX para desarrollar las aplicaciones.

Cuando usarla:

* Si se desean construir aplicaciones con alto rendimiento de modelos de data mining embebidos que utilizan entornos con multiprocesadores.
* Si se quiere tener un absoluto control sobre todos los elementos de los procesos de construcción de modelos.
* Si se necesitan combinar operaciones y tecnicas de data mining alternativas en aplicaciones complejas.
* Si se quiere trabajar con una solución que puede comunicar una aplicación data minino para enlazar con sus necesidades.

## **Pilot Discovery Server de Pilot Software, Cambridge MA, USA.**

Puntos Clave:

* Herramienta para clasificación y predicción.
* Basada en la tecnología de árboles de decisión CART.
* Trabaja sobre UNIX y Windows NT

Ventajas:

* Buena representación del análisis de resultados
* Es fácil de usar y de entender.
* Muy integrada con sistemas gestores de bases de datos relacionales.

Aspectos a tener en cuenta:

* Solamente indicada para clientes de los programas para soporte a la toma de decisiones de Pilot.
* Solamente cubre un especifico sector del espectro del data mining.
* Sólo trabaja con datos almacenados en bases de datos relacionales.

Cuando usarla:

* Si se desea optimizar las campañas de marketing.
* Si se necesita interpretar fácilmente los resultados sin realizar un gran refinamiento de los modelos.
* Solo si se están utilizando los programas para soporte a la toma de decisiones de Pilot.
* No está indicada si se quieren resolver los problemas utilizando diferentes técnicas.

## **SAS Solution for Data Mining de SAS Institute, Cary, NC, USA**

Puntos Clave:

* Un gran número de herramientas de selección, exploración y análisis de datos para entornos cliente-servidor.
* Las opciones de data mining incluyen: aplicaciones de redes de neuronas, de árboles de decisión y herramientas de estadística.
* Aplicaciones portables para un gran número de entornos PC, UNIX y mainframes.

Ventajas:

* SAS ofrece data warehouse y análisis de datos.
* Conjuntos extensibles de herramientas de manipulación y visualización de datos.
* SAS tiene una gran experiencia en herramientas estadísticas y de análisis de datos.

Aspectos a tener en cuenta:

* La oferta para hacer data mining es una mezcolanza de todas las técnicas SAS existentes.
* Integración con la programación en 4GL.
* No soporta el análisis de asociaciones.

Cuando usarla:

* Si ya se utiliza SAS para almacenar, administrar y analizar los datos.
* Si se va a utilizar SAS para la construcción del data warehouse.
* Si es necesaria una alta funcionalidad en la manipulación de datos.
* Si se es experto en estadística y se quieren utilizar las funciones estadísticas de SAS.

## **MineSet, de Silicon Graphics, Mountain View, CA, USA**

Puntos Clave:

* Paquete de herramientas para Data mining y visualización.
* Proporciona algoritmos para la generación de reglas para clasificación y asociaciones.
* Trabaja sobre plataformas SGI bajo IRIS.

Ventajas:

* Ofrece herramientas de visualización para los datos y los modelos generados.
* Suporta muchas operaciones de data mining.
* El gestor de herramientas actúa como un punto central de control y permite el acceso y transformación de los datos.

Aspectos a considerar:

* Requiere un servidor SGI.
* La gran cantidad de opciones y parámetros puede provocar confusión en usuarios noveles.
* Las herramientas de visualización necesitan mucha preparación y personalización de los datos para producir buenos resultados.

Cuando usarla:

* Si se quieren detectar patrones por visualización.
* Si se quieren construir aplicaciones que representen los resultados de data mining a través de visualización.
* Si se dispone de equipos de Silicon Graphics
* Esta indicada para VARs que quieran desarrollar soluciones personalizadas de data mining usando MineSet.

## **SPSS, de SPSS, Chicago IL, USA**

Puntos Clave:

* Herramientas de escritorio para clasificación y predicción, clustering, y un gran rango de operaciones estadísticas.
* Proporciona una herramienta de redes de neuronas además de productos de análisis estadístico.
* SPSS para Windows y Neural Connection son productos que trabajan en modo monopuesto en plataformas Windows.

Ventajas:

* Las funciones de análisis estadístico complejo son accesibles a través de una interface de usuario muy bien diseñada.
* Neural Connection ofrece un amplio rango de opciones y funciones a través un entorno de desarrollo muy fácil de usar.
* El lenguaje de scripts permite una gran personalización del entorno y el desarrollo de aplicaciones estadísticas aisladas.

Aspectos a considerar:

* Para analistas de datos y estadísticos, más que para usuarios finales.
* SPSS CHAID carece de la funcionalidad de otros productos de escritorio de árboles de decisión. •
* Neural Connection es un producto aislado: la base de la integración con SPSS es a través de transferencia de datos, que se limita a la importación de 32.000 registros.

Cuando usarla:

* Si se necesita un análisis complejo combinando estadística con árboles de decisión y redes de neuronas.
* Si se disponen de grandes conocimientos estadísticos y se quiere utilizar data mining basado en IA.
* Si se necesita verificación estadística de los resultados encontrados.
* Si es preciso construir aplicaciones de análisis departamental para escritorio.
* Si tiene un presupuesto ajustado.

## **Syllogic Data Mining Tool, de Syllogic, Houten, The Netherlands**

Puntos Clave:

* Herramienta con entorno de trabajo multi-estratégico con interface visual.
* Soporta análisis de árboles de decisión, clasificación k-vecino más próximo, y análisis de clustering y asociaciones por k-means.
* Trabaja sobre Windows NT y en estaciones UNIX con uno o varios procesadores

Ventajas:

* La interface visual permite a los usuarios construir proyectos de data mining enlazando objetos.
* La versión está optimizada para entornos masivamente paralelos y validos para grandes bases de datos.
* La empresa también ofrece un gran número de servicios de consultaría en las áreas de datawarehousing y data mining.

Aspectos a considerar:

* La interface y la presentación de resultados necesita algunos refinamientos para ser utilizada por usuarios finales.
* DMT/MP no soportan el mismo rango de operaciones que DMT

Cuando usarla:

* Si se necesita servicio de consultoría a la vez que se desarrolla el proyecto de data mining con un entorno de datawarehousing.
* Si se necesita utilizar gran número de operaciones de data mining.
* Si se quiere utilizar una herramienta similar en el escritorio y en el entorno MP.

## **Darwin de Thinking Machines, Bedford MA, USA**

Puntos Clave:

* Herramientas de desarrollo de data mining de tipo cliente-servidor para la construcción de modelos de clasificación y predicción.
* La construcción de modelos utiliza algoritmos de redes de neuronas, árboles de inducción y k-vecino más próximo.
* Trabaja sobre plataformas Sun de Solaris, AIX de IBM y SP2, con clientes Motif. También existen versiones cliente que trabajan sobre Windows.

Ventajas:

* Ofrecen buena cobertura al proceso completo de descubrimiento del conocimiento.
* Pone el énfasis en el desarrollo de modelos predictivos de alto rendimiento.
* Proporciona escalabilidad para soportar paralelización.

Aspectos a considerar:

* Mejor para analistas de datos y desarrolladores de aplicaciones que para los usuarios de negocio.
* Es preciso familiarizarse con las diferentes opciones de Darwin para cada tipo de modelo si se quiere obtener el mejor resultado de la herramienta.
* No soporta análisis no supervisado de clustering o de asociaciones.

Cuando usarla:

* En la construcción de aplicaciones de data mining para gestión de relaciones entre clientes.
* Si se necesita una herramienta que ponga mucho énfasis en modelado por clasificación y predictivos.
* Si se dispone de una gran compleja base de datos que precise la potencia de una plataforma con multiprocesadores.
* Si se necesita observar la creación de los modelos de data mining, Darwin proporciona múltiples algoritmos y varias opciones de refinamiento.
* Si se quiere usar las herramientas de data mining para auxiliar la gestión de redes Thinking Machina tiene objetivos muy explícitos en este sector y ya colabora con Cabletron.