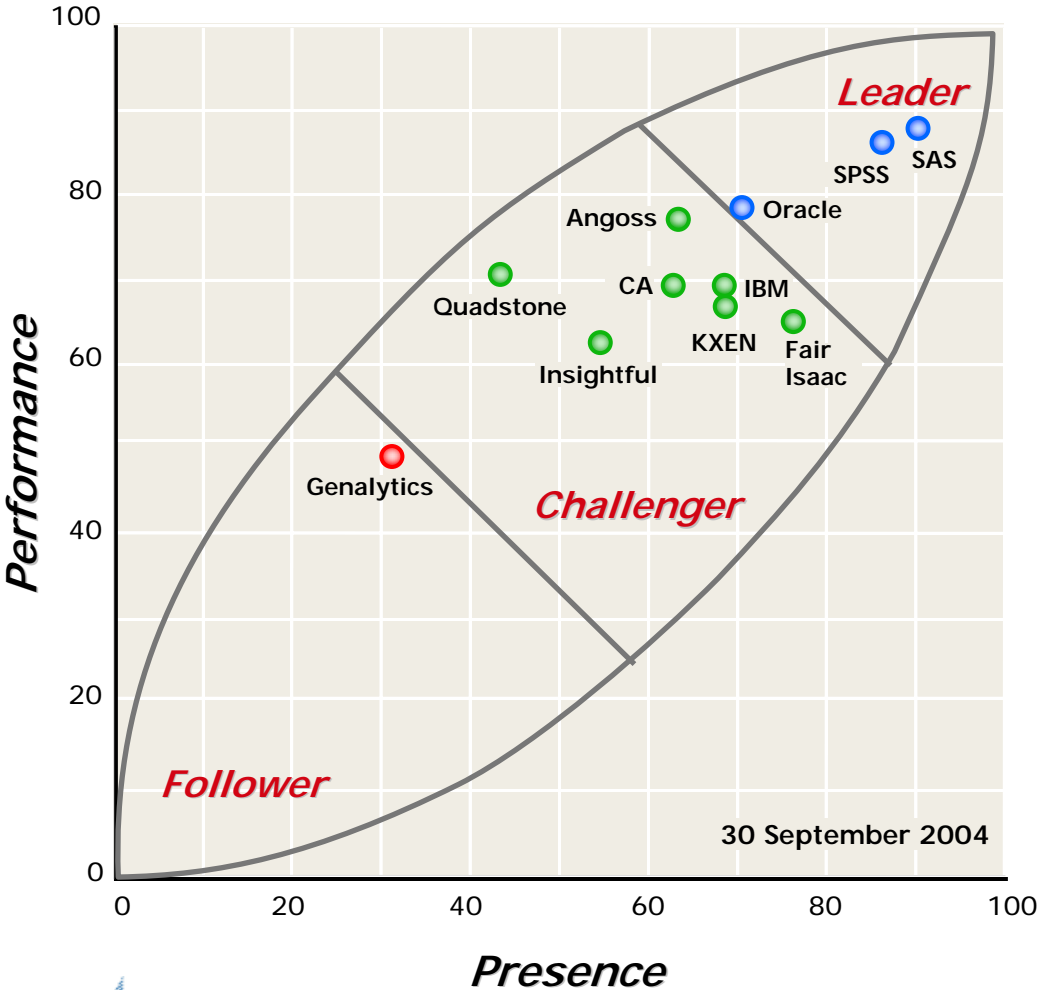




# Data Mining Tools

METAspectrum<sup>SM</sup> Evaluation



30 September 2004



META Group is a trademark, and METAspectrum is a service mark, of META Group, Inc.  
Copyright © 2004 META Group, Inc. All rights reserved.

## Market Definition

The data mining market consists of software vendors offering tools that extract predictive information from large data stores, which can then be analyzed to enhance corporate data resources and generate predictions regarding business trends and behavior. Specifically, these tools provide statistical data models (classification or clustering studies, linear regression, and current or predictive modeling) and utilize visualization functions to support the analysis of massive quantities of data stored by customer organizations. Data mining tools may be implemented on existing customer platforms or integrated with other applications as part of a larger data quality initiative or business intelligence (BI) strategy. Data mining tools provide both developers and business users with an interface for discovering, manipulating, and analyzing corporate data.

## Market Forecast

We expect the data mining market to expand 10%-20% annually over the next few years, with services growing at a commensurate pace. The market is crowded, which will keep some vendors from growing rapidly. Specialty, niche-based offerings will continue to find their place without threatening established market leaders, though this will not last forever. Consolidation within the data mining industry looms in the near future, though we do not believe it will be widespread in 2004/05. This is particularly significant for the younger, smaller, and niche-focused vendors, which could emerge as attractive acquisition targets for larger competitors and leading data quality vendors (and by extension, leading ETL or BI vendors), many of which already exist as key technology partners. New entrants to the data mining market will find it crowded with experienced analytics providers, including many with impressive customer bases already. Without long-term vision and substantial backing, it will be difficult for new players to compete.

Going forward, we expect significant impact to come not entirely from data mining vendors themselves, but from business application vendors (e.g., SAP, Siebel) that are embedding data mining technology from these vendors. Embedding data mining directly into the business process enables recommendations and predictions to be provided directly in the GUI of the application. Immediate value, however, will be limited to data mining functionalities/initiatives that can be easily understood by non-specialized end users. The overall data mining market may be splitting into two segments: 1) advanced or pure data mining, and 2) data mining for a mass audience of less-technical business users. The latter technology is the one being embedded into enterprise applications and/or BI platforms.

The following vendors and products were evaluated:

- Angoss Software KnowledgeStudio 4.2 and Mining Manager 2.1
- Computer Associates CleverPath Predictive Analysis Server 3.0
- Fair Isaac Enterprise Decision Management suite
- Genalytics Predictive Suite 5.0
- IBM DB2 Intelligent Miner
- Insightful Miner 3.0
- KXEN Analytic Framework 3.0
- Oracle Data Mining
- Quadstone System V. 5
- SAS Enterprise Miner 5.1
- SPSS Clementine 8.5

## Key Findings

The data mining market comprises large enterprise software vendors (e.g., Oracle, SAS, IBM, Computer Associates) and smaller data analysis specialists (e.g., SPSS, KXEN, Angoss). In addition, there are a few smaller, up-and-coming players (Quadstone, Genalytics) that choose to focus on specialty environments in light of smaller market shares. Most vendors evaluated in this METASpectrum, however, offer comprehensive data mining functionality encompassing highly scalable architectures supporting flexible workbenches and user interfaces employing various data mining models and algorithms. Market presence criteria are important to the vendor selection process, since not all vendors possess the same data mining vision or marketing strategy. In many cases, partner programs are required to assist in implementation, and are key in driving international

revenues as well. Business drivers vary according to history and experience, as some vendors have deep roots in the data analysis market while others are still learning how to distinguish their solutions from those of competitors with larger market shares. Performance criteria are heavily weighted toward technology and product functionality, though potential buyers should also pay close attention to pricing, services/support, execution (in keeping with market trends and customer demands), and financial viability.

### Leaders

Leaders in this market have stable, mature products that excel in nearly all aspects of data mining functionality. Moreover, the leaders in this space have large market shares relative to the other players. In some cases, their technical capabilities and overall functionality are not overwhelmingly superior to those of challenging vendors, but in a specialized market such as data mining, only a small number of vendors can excel to this degree in both presence and performance. In addition to market share, leaders are also distinguished by the breadth of their marketing programs, geographic coverage, technological investments, and commitment to quality implementation and support services.

### Challengers

Challengers in this market are primarily characterized by a slightly narrower scope of data mining functionality and/or less commitment to the industry in general, versus the leaders. Many of these vendors are large software houses that offer various software solutions spanning multiple IT fields/markets. Specializing in the data mining industry is not a priority for these organizations, though the functionality their products provide is more than sufficient for most data mining implementations. Other challengers possess the commitment to data mining and analysis, but do not yet have the resources or customer bases to drive market share and excel in other presence and performance criteria.

### Followers

The only follower evaluated in this METASpectrum is Genalytics, a relative newcomer to the data mining market. Founded in 1998, Genalytics is a small organization with few customer installations, nearly all of which are in the financial services industry. The Genalytics approach to data mining focuses exclusively on applying genetic algorithms to predictive analytics. The company's vision for data mining calls for a focus on product development (with key emphasis on becoming the leader in genetic algorithms) in the short term, with the goal of competing for significant market share in advanced analytics within three to five years. It may not be entirely fair to label Genalytics a "follower" in a market where it clearly and purposefully pursues an alternative data mining strategy, though it is important for prospective customers to understand the distinction between the Genalytics approach and that of traditional data mining vendors and market leaders.

## Bottom Line

The data mining market will increasingly intersect with related solutions in data quality and data integration (and by extension, business intelligence and business performance management), but we do not see these vendors being absorbed into these markets. Inevitably, there will be some consolidation in the market, but the variety of specialized data mining solutions should always guarantee competition. As data mining evolves, however, we expect vendors to become more heavily invested in predictive analytics, resulting in stagnation and/or pricing degradation among other vendors in the market. Technology remains a critical differentiator when evaluating products, but potential buyers should not become fixated solely on functionality. Presence criteria areas are equally important in this market.

**Business Impact:** Data mining tools should be considered part of an organization's IT portfolio for driving new business opportunities and making informed business decisions, due to their predictive analytic capabilities. Data mining tools are also proving themselves valuable in cost avoidance or stop-loss equations.

## Vendor/Product Evaluation

| Vendor Contact Information   |   |
|------------------------------|---|
| <b>Vendor:</b>               | SPSS Inc.   |
| <b>Headquarters Address:</b> | 233 S. Wacker Drive, 11th floor<br>Chicago, IL 60606-6307 |
| <b>Phone:</b>                | (312) 651-3000  |
| <b>URL:</b>                  | www.spss.com  |

### **Evaluation Summary**

SPSS remains a leader in the current data mining market and has no significant weakness or disadvantage for prospective customers. The Data Mining Workbench is the user interface in Clementine 8.5 for manipulating data, designing workflows (or “streams”), managing nodes, and drilling into data to obtain specific information. SPSS continues to drive superb market positioning and share, and offers users leading benefits related to marketing awareness/industry reputation, geographic coverage, support services, and business drivers, such as core competencies.

### **Strengths**

- ▶ Clementine offers enhanced data exploration features through support of visualizations (Web graphs) showing associations for wide data sets, including large numbers of variables.
- ▶ A rich set of data preparation functions is provided out of the box, with little coding required to collect data specifically for modeling purposes. These features further enhance interactive visualization by enabling automatic calculation of values on the fly.
- ▶ The “publisher” node packages everything needed to deploy streams/results into an operational system. The “predictive applications wizard” takes users through a six-step process for exporting streams into real-time environments.
- ▶ Clementine offers internal text mining capabilities (thanks to the SPSS’s acquisition of Lexiquest technology), making it possible to extract information from unstructured data. The text mining node in Clementine identifies/pulls key concepts to be analyzed and/or combined with other customer data models or concepts.
- ▶ Using Web importer technology acquired from NetGenesis, Clementine can incorporate Web data into the modeling environment and process Web behavior through Web logs.
- ▶ Clementine is designed to fit within the SPSS Enterprise Analytic Platform, which provides a framework for using Clementine in real-time production environments. The framework also enables seamless movement between predictive analytics and interaction optimization components.

### **Limitations**

- ▶ Clementine supports a large number of data mining models and algorithms, but does not yet offer support for genetic algorithms.
- ▶ SPSS does not offer ASP or hosted services for data mining, though some customers have taken results through SPSS’s Web deployment tools.
- ▶ Clementine does not provide its own portal interface, though users can work with deployed models through Cleo, a Web interface that can be embedded in a portal.