Technical report

Data Mining Deployment for High-ROI Predictive Analytics

Four decision optimization strategies based on successful deployments and the CRISP-DM methodology

Table of contents

Executive summary 2
Data mining deployment ROI requirements 4
Strategy 1—Rapid-update batch 6
Quantifying automated deployment cost savings 8
Strategy 2—Extreme-volume batch 9
Strategy 3—Packaged real-time 11
Strategy 4—Customizable real-time 12
Managing predictive models to achieve ROI and compliance goals 14
Decision optimization strategy next steps 15
About SPSS Inc. 15

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © 2005 SPSS Inc. All rights reserved. DMDWP-0505

Executive summary

Data mining is evolving quickly into a mainstream, must-have technology that puts predictive insight in the hands of decision makers. This evolution is driven by the tangible return on investment (ROI) that organizations in every industry are realizing, for example:

- 100 percent improvement in direct mail campaign response rates
- 25 percent decrease in customer churn
- $1.8 billion revenue increase by prioritizing government collections

Decision optimization—the key to high returns

While the potential business value of predictive analytics is clear, processes for realizing that value—strategically deploying data mining results to create predictive analytics solutions—are often less clear. Many organizations are successfully applying data mining to discover new insights, closing what Gartner has dubbed the knowledge gap (figure 1), but then failing to execute decisions based on those insights to achieve their desired returns.

Knowledge is not the problem

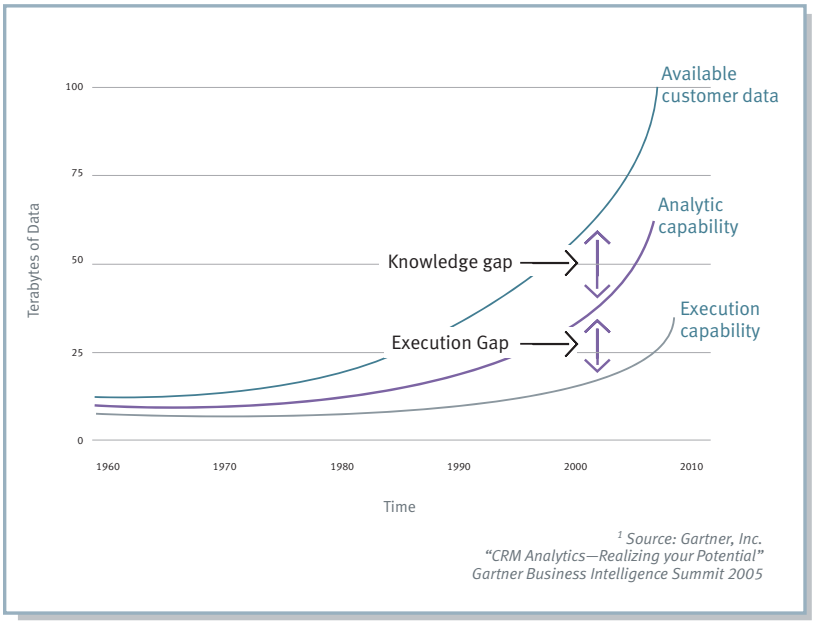


Figure 1: Organizations struggle to both analyze increasing data volumes and execute decisions based on the analyses

According to Gareth Herschel, Research Director, Gartner Inc., who discussed this at Gartner Business Intelligence Summit 2005, "The growing number of tools and implementations have begun to seriously address the issue of the knowledge gap, but generally ignored the problem of the execution gap. Most enterprises take a top-down approach that looks at the available data and analyzes it, but then lack the ability to effectively deploy the results of the analysis. A bottom-up approach to analytics is needed that begins by identifying the desired decision, and works back to the analysis that will drive the decision and the data that will drive the analysis."[1]

The major failure of most data mining implementations is that they are "siloed"—tactically focused on analysis. High-ROI predictive analytics requires an approach that's strategically focused on decision optimization—deploying the results of advanced analysis to achieve business objectives. The CRoss Industry Standard Process for Data Mining (CRISP-DM) is the de facto standard for implementing data mining as a strategic business process to optimize enterprise decision making.

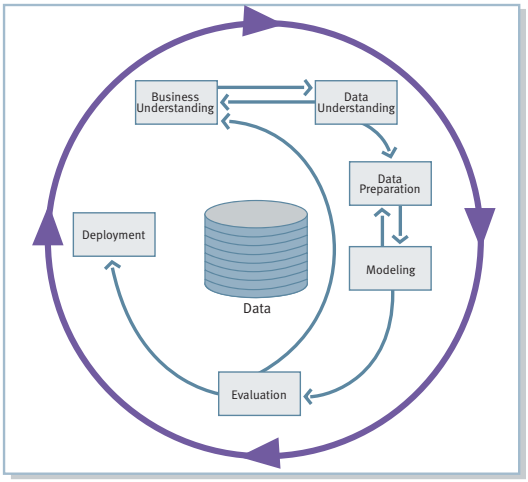


Figure 2: CRISP-DM focuses data mining on results deployment to optimize decisions

CRISP-DM is an open standard developed by an international consortium of over 200 organizations, including DaimlerChrysler, NCR, and SPSS, to apply data mining to a wide range of business objectives. The process begins with an understanding of your business goals and available data, and ends with the deployment of data mining results into business operations to optimize critical decisions and deliver measurable ROI (figure 2). For example, a CRISP-DM data mining initiative to increase customer retention by 25 percent deploys churn reduction models into call center applications to optimize recommendations made by customer service representatives. These models are deployed with supporting data mining tasks such as those for data preparation and post-processing.

Four strategies based on CRISP-DM and successful implementations

This white paper presents four decision optimization strategies, based on the proven CRISP-DM methodology, to help you deploy the results of advanced analysis and plan a predictive analytics solution for the greatest possible returns. These strategies are generalized from real-world, successful implementations using the Clementine® data mining workbench's flexible range of deployment options—the widest available.

Decision optimization strategies discussed in this paper:

1. **Rapid-update batch**—deploy data mining using cost-efficient, high-speed batch scoring capabilities to focus on updating scores quickly and easily
2. **Extreme-volume batch**—deploy data mining with the strategic use of coding, making coding cost trade-offs, to focus on high-speed scoring that scales to extreme data volumes
3. **Packaged real-time**—deploy data mining into packaged applications that minimize integration risk and focus on high-speed, real-time scoring at specific customer touchpoints
4. **Customizable real-time**—deploy data mining into customizable applications to focus on real-time scoring for different roles within an organization or unique business goals for which packaged applications do not exist

Predictive analytics

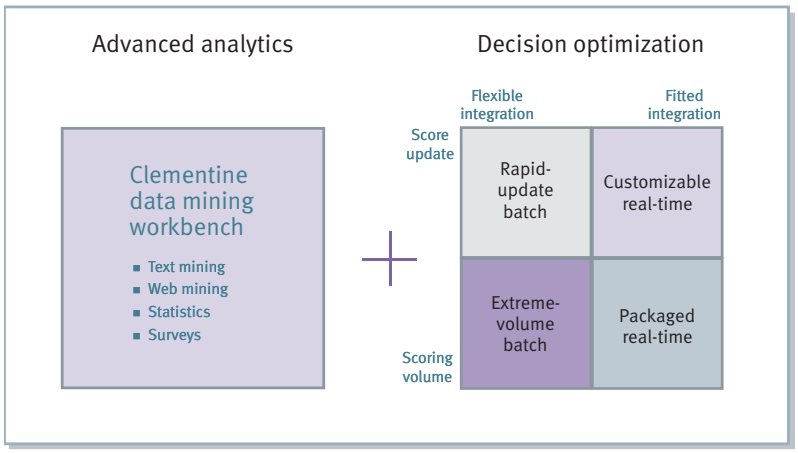


Figure 3: Decision optimization strategy matrix

Generalized from the most common decision optimization objectives, these four strategies are described in the context of two key deployment considerations—scoring focus and integration focus—where difficult trade-offs may have to be made (figure 3). Deployment issues often exist around making scoring focus trade-offs between the ease of score updates and scoring volume—and deciding which objective is a priority when "operationalizing" data mining results. Integration focus trade-offs often need to be made between the benefits of tight integration within a specific interaction channel, such as the Web or your call center, and greater integration flexibility. SPSS predictive analytics experts are available to help guide you through these trade-offs and tailor strategies to fit your specific business objective, processes, and IT systems.

Data mining deployment ROI requirements

Predictive analytics solutions deliver some of the most quantifiable returns of any technology investment. Yet the economics of data mining—from model development to deployment—must be understood and managed effectively to achieve those returns. Before reviewing the four decision optimization strategies presented in this paper, it's important to understand these economics and how Clementine data mining helps you maximize returns.

Data mining deployment for high-ROI predictive analytics requires:

- An open architecture for ease of integration
- Rapid model development
- Flexible deployment options

Open architecture capitalizes on existing systems

To maximize returns, models must be developed quickly and deployed cost-effectively for use within current operational systems and business processes. Clementine's open architecture capitalizes on your existing IT investments with a complete in-database mining platform designed to scale to enterprise demands. This standards-based data mining platform provides integrated capabilities for the entire data mining process, including flexible model deployment options for nearly every operating environment and auditable model management. Data mining platforms from other vendors with highly proprietary technology add unnecessary risk and costs—mandating additional expenses for software, hardware, and services that lessen your chances of realizing a significant return.

Rapid modeling accelerates time-to-value

Soft costs, such as the time cost of a delay in deploying data mining results, should also be factored into data mining ROI calculations. For example, what is the time cost of a 30-day delay in deploying a customer retention model that is predicted to save 2,000 customers per month? Lifetime value calculations provide a concrete time cost for such a scenario. Other significant time cost scenarios include businesses that operate in volatile markets—where delays in model deployment may mean that the model is outdated shortly after it's deployed.

Clementine's visual workflow interface for rapid predictive modeling generates cost savings by minimizing data mining time-to-value, the time required to develop and deploy models that deliver value within business operations. Clementine represents the entire CRISP-DM process as a "stream" that maps process steps as you work (figure 4)—from open access to databases and preparing multiple data sources, to in-database modeling and data mining deployment. The Clementine stream approach to data mining makes it possible to create predictive models quickly and deploy them in an integrated manner with other critical predictive analytic functions represented in the stream, such as Web mining or statistical analysis.

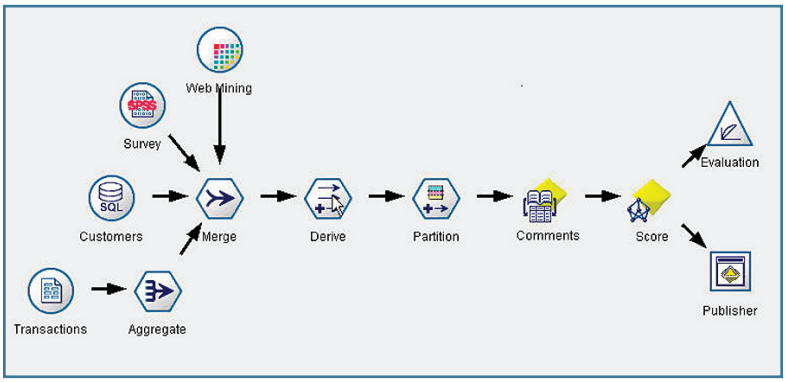


Figure 4: Clementine streams are maps of the entire CRISP-DM data mining process

Flexible options minimize deployment costs

Different operating environments have different data mining integration requirements. Clementine's flexible deployment options provide you with the ability to integrate data mining within almost any operating environment cost-effectively. Most Clementine deployment options deploy the entire CRISP-DM data mining process, automating complex pre- and post-processing tasks, such as combining different data sources and ranking by score, to eliminate hand-coding costs that would reduce ROI. Exporting a complete data mining process—rather than just scoring code—helps you realize significant cost savings over the life of the deployment by eliminating programming requirements. The programming costs related to other vendors' data mining deployment options, that deploy scoring code that must be integrated within existing systems by hand, can quickly add up to decrease returns.

Strategy 1—Rapid-update batch

Rapid-update batch scoring is one of the most widely used Clementine decision optimization strategies—appropriate for operating environments that don't require real-time responsiveness. For example, many companies use rapid-update batch scoring to keep customer databases updated—providing decision makers with the latest predictive insight via customer relationship management applications. This strategy and its related deployment options provide you with the flexibility to integrate data mining within a wide range of operating environments.

Clementine's batch mode feature and Clementine Solution Publisher were developed to help you address the challenge of cost-effective data mining deployment and rapid updating. Deployed Clementine data mining streams run in the background of business operations without the need for the Clementine client interface. Clementine batch mode is executed from a command line, while Clementine Solution Publisher, a flexible scoring component, can be embedded within an application. These deployment options can be scheduled to update scores within a database on a regular basis, as the application requires—monthly, weekly, daily, or even hourly—using the latest data. Since both of these options execute entire Clementine streams, updating scores simply requires rerunning the stream. Rapid-update batch scoring with Clementine batch mode or Clementine Solution Publisher can even be fully automated using the Predictive Enterprise Manager module of SPSS Predictive Enterprise Services™ (see page 14).

Unlike other data mining solutions, Clementine deploys the entire data mining process, including critical data preparation, modeling, and model scoring tasks, for use within leading databases such as IBM® DB2® , Oracle® Database, and Microsoft® SQL Server™. These Clementine deployment options ensure high-performance in-database scoring using a three-tiered architecture (figure 5) that takes advantage of your database's indexing, optimization, and in-database mining functionality. Generally, most companies only use the Clementine client to perform batch-scoring operations on an ad hoc basis, and run scheduled batch scoring operations using Clementine batch mode or Clementine Solution Publisher. The Clementine client passes stream description language (SDL), which describes the data mining tasks that need to be performed, to Clementine Server. Clementine Server then analyzes these tasks to determine which it can execute more optimally in the database—minimizing data movement. After the database runs these functions, the remaining and often aggregated data are passed to Clementine Server.

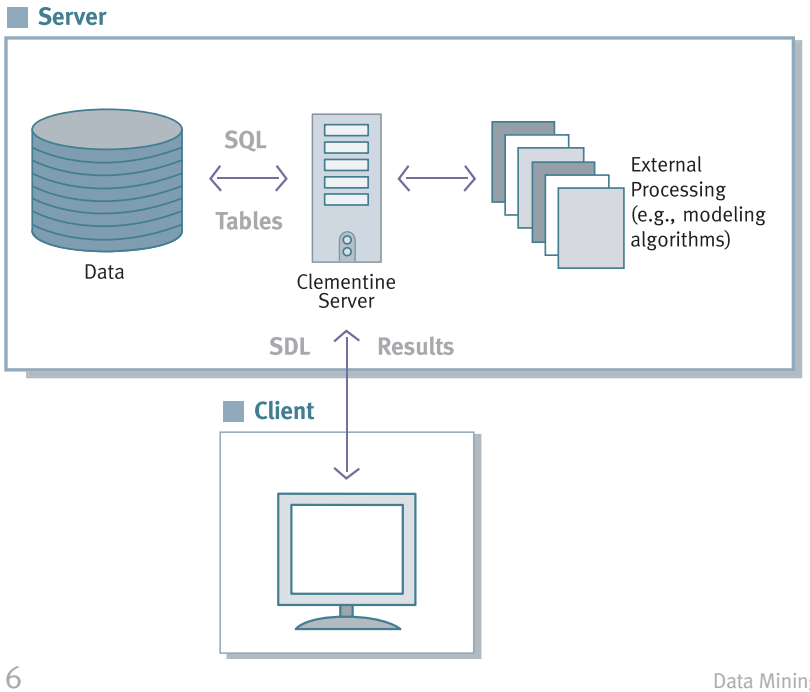


Figure 5: Clementine scales in-database batch scoring using a three-tiered architecture



Capitalizing on databases to generate £33 million

Standard Life, one of the world's leading mutual financial services companies, applied rapid-update batch scoring to capitalize on customer databases throughout its organization. Standard Life Bank, a division of Standard Life, launched its Freestyle Mortgage product with extensive television advertising and public relations campaigns. Soon thereafter a number of similar products appeared from rival providers—making it essential that Standard Life Bank both consolidate and continue to expand its share of the mortgage market. The bank turned to Clementine data mining to model the key attributes of valuable customers most likely to be attracted to mortgage offers to improve cross-selling within the Standard Life group.

Donald MacDonald, customer data analyst at Standard Life further explains, "Our vision was to increase both the speed at which we build our models and the sophistication of those same models. This ultimately leads to improved customer communications as well as greater returns on the bottom line." Standard Life Bank analysts used Clementine to develop a propensity model for a mortgage offer identifying customers most likely to respond, and used it to score Standard Life group's databases and prospect pool. This enabled the bank to focus a direct mail campaign on the best prospects—resulting in a 9x improvement in response rate and £33 million (approx. $47 million) in mortgage application revenue.

HEW taps Oracle database to increase customer focus

Hamburgische Electricitäts-Werke AG (HEW) chose Clementine data mining for its unique ability to tap into the value of its Oracle database. The deregulation of the German electricity market in 1998 caused a shake-up in the energy sector. Stripped of their position as a bureaucratic monopoly, public electricity suppliers suddenly faced competition characterized by comparatively low prices.

HEW adopted data mining as the core element of its new marketing approach—implementing automated customer segmentation as the basis for more targeted customer care. Clementine enabled HEW to create accurate customer segments by directly tapping into their Oracle data warehouse, which contained customer questionnaires, power consumption rate measurements, and a continual flood of other new customer data. The Clementine data mining solution was "simple to interpret and easy to understand for each of HEW's various hierarchical levels," said Peter Goos, of HEW's marketing and sales management department.

In addition to the ability to access, prepare, and score data within Oracle databases, recent enhancements to Oracle integration with Clementine enable you to perform in-database modeling with Oracle Data Mining. Data mining workbenches from other vendors force you to move data into their proprietary formats—adding unnecessary steps that significantly delay modeling progress and impede "train-of-thought" data mining.

Quantifying automated deployment cost savings

Estimating true deployment costs and potential savings are important to arriving at accurate ROI projections and comparisons. While assumptions and trade-offs vary based on your specific business situation, maintaining hand-coded data mining deployment always adds significant costs. But in certain extreme-volume scoring situations, these coding costs may be a necessary and acceptable trade-off to achieve the highest possible performance (see *Strategy 2—Extreme-volume batch on page 9*).

How much would your organization save if you could automate data mining deployment? To quantify the estimated potential savings of automated deployment vs. hand-coded deployment, consider the following factors:

- **Programming costs**. How much does your organization spend when programmers hand code data mining steps? What are the personnel costs for your programming staff, including overhead and QA costs?
- **Lines of code per day**. How many lines of bug-free code can your programmers code per day?
- **Initial deployment costs**. Keep in mind that embedding data mining within a business application typically requires 3,000 lines of code.
- **Redeployment costs**. Frequent updates must be made to at least 20 percent of the code.

To illustrate the costs of a hand-coded data mining deployment, consider the case of a financial services company that deployed a data mining process for improving customer retention and detecting fraud into their call center application. The company's programmers produce 10 lines of bug-free code per day. For 220 days of work, the company spends $100,000 for programming personnel costs, or about $450 per day. The company redeploys the model annually, at a cost of $27,000 for 60 programming days. The total cost of the company's initial deployment is $100,000. Annual model updates add another $27,000 per year. Annual maintenance of the hand-coded model deployment causes costs to skyrocket over a five-year period. The financial services company's hand-coded data mining deployment totals $208,000 over the five-year life of the deployment (figure 6).

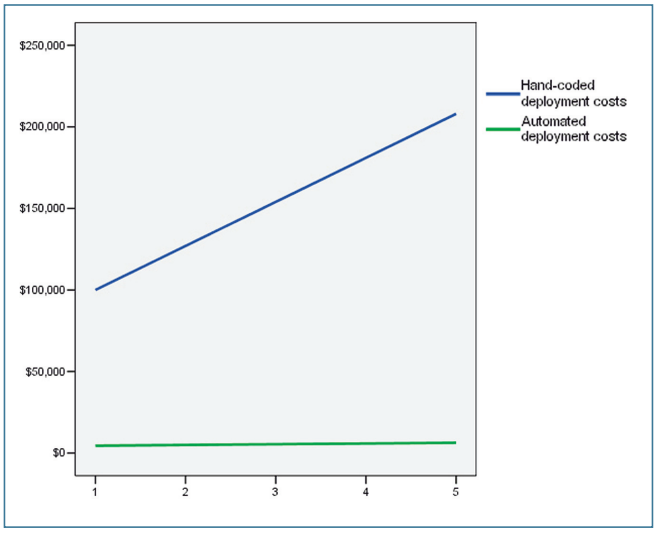


Figure 6: Hand-coded deployment costs skyrocket vs. automated deployment

Automated deployment costs remain predictably low. With automated deployment, initial deployment costs are minimized to $4,500 for 10 days testing and design work. Annual redeployment costs are minimized to $450 for one day of testing and implementation work. Over the same five-year period, automated deployment results in a cost savings of $201,700 versus hand-coded deployment.

Strategy 2—Extreme-volume batch

"Extreme" batch scoring for high-volume operating environments necessitates the strategic use of hand coding—requiring you to make careful trade-offs to determine when coding costs are justifiable to produce the best economic return. Coding data preparation, or even the entire data mining process, dramatically improves performance in extreme situations in which billions of scores need to be generated within a short processing window. Custom SPSS high-speed batch scoring engines can be deployed to generate more than 20 billion scores in under an hour.

Most Clementine deployment options deploy Clementine streams into an "interpretive" scoring engine that eliminates coding requirements to minimize costs and data mining time-to-value (figure 7). The data mining process is described in a file that must be interpreted by a scoring component or application. Extreme-volume batch scoring applies a "compiled" approach in which at least some of the data mining process is coded in a programming language and compiled to make the code readable by a computer. By definition, a compiled approach is faster than an interpretive approach.

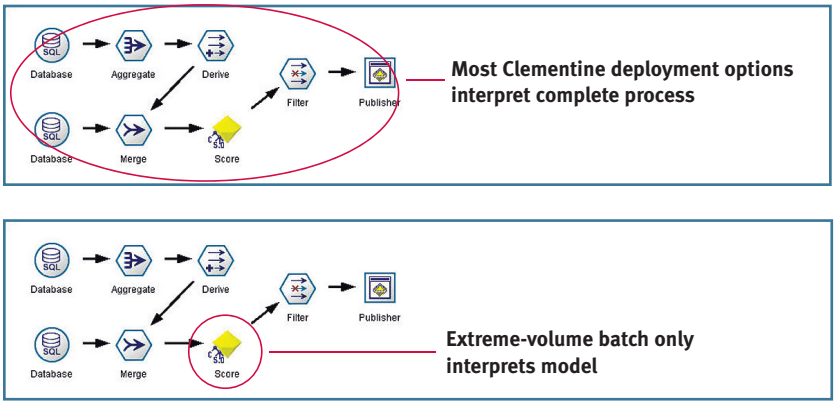


Figure 7: Extreme-volume batch most often uses PMML model deployment with hand-coded data preparation

The performance bottleneck in extreme scoring environments is most often in the data preparation phase rather than the raw scoring, so a hybrid approach is most often used by SPSS high-speed scoring options. Clementine exports predictive models into high-speed scoring engines using Predictive Model Markup Language (PMML), the industry-standard XML mark-up language for data mining models. The PMML standard, a vendor-neutral method of exchanging models, is maintained by the Data Mining Group, an independent organization led by major data mining and database vendors. PMML model deployment is partially an interpretative approach to scoring—models are described in PMML files and interpreted—while data preparation is hand coded and compiled to scale to high-volume demands. Custom batch scoring engines based on SPSS PMML-scoring components such as SmartScore®, a software development kit, are available to scale to extremely high-volume scoring requirements. Clementine PMML models can also be deployed for in-database scoring using IBM DB2 Scoring, improving performance by eliminating the need to remove data from DB2 databases.



Twenty billion scores in less than an hour with grid computing

The marketing services division of a global information solutions provider uses a combination of Clementine batch mode and an SPSS high-speed scoring engine to provide a data mining service to direct mail retailers that improves customer targeting. The retailers send their customer data to the solution provider for inclusion in a "cooperative" database, enabling customers to draw value from the entire set of data instead of just the subset they contribute. Scores for each customer in the database are generated on a monthly basis using the entire cooperative customer dataset. This results in hundreds of models, which are run against the entire customer dataset to generate hundreds of scores for each customer record. The custom scoring solution quickly prepares and scores 270 million rows of customer records with 200 attributes using 800 models—20 billion scores are generated in under an hour, on cost-efficient hardware. This same scoring operation took the previous data mining vendor multiple days and a mainframe computer to perform within a rigid, expensive, proprietary process.

The company's custom scoring solution scales using a grid-computing configuration of 16 Linux-based PCs—at a cost of less than $10,000. Clementine's open architecture enables organizations to utilize existing computing resources, rather than purchasing additional hardware, to keep the total cost of ownership as low as possible. The hardware costs of custom scoring solutions can even be eliminated completely by using grid computing that taps into unused processing power throughout corporate networks. If required, Clementine custom batch scoring solutions can be scaled to meet increasing customer requirements by simply adding additional processing power to the computing grid.

Generating $1.8 billion using government legacy systems

Booz Allen Hamilton, a global strategy and technology firm, was named SPSS partner of the year for its work assisting federal agencies in improving business results and increasing revenue. Booz Allen Hamilton, which was rated the #1 consulting firm in a 2003 customer survey, used Clementine data mining to assist a federal government agency that was facing an increasing backlog of collections.

With Clementine, Booz Allen Hamilton developed predictive models to recommend a collection priority scheme that optimized the agency's limited resources and aligned its operations with new strategic objectives. Within 18 months of project commencement, a joint Booz Allen/client team deployed these models into a legacy systems environment by converting them into XML and C code, and delivered measurable business results within 24 months. The project resulted in an estimated $1.8 billion in annual revenue, representing a 600:1 first-year return on investment for the agency.

"Government agencies constantly have to make intelligent choices in order to fulfill their mission. This requires them to apply limited resources to increasing workloads while increasing customer service levels. Our work with SPSS has demonstrated that predictive analytics can be an effective method of providing clients with an immediate improvement in business results, even in legacy system environments," said Jill McCracken, Ph.D., Booz Allen associate, during the partner of the year award ceremony.

Strategy 3—Packaged real-time

Applications of predictive analytics for certain business objectives and operating environments, such as cross-selling in a call center or Web environment, demand real-time, high-speed scoring that scales to high volumes. Companies applying predictive analytics to customer interaction channels like these—where data from each interaction must be immediately analyzed during the interaction to optimize offers—are increasingly turning to packaged applications to minimize integration risk and time-to-value.

SPSS predictive analytics applications such as PredictiveCallCenter™ and PredictiveWebSite™ are designed to capitalize on existing customer interaction points and CRM systems. A "packaged real-time" decision optimization strategy using Clementine's PredictiveCallCenter deployment option focuses on rapid, real-time execution for hundreds of customer service representatives. PredictiveCallCenter integrates with call center systems to instantly determine and recommend the optimal agent-customer interaction, such as an up-sell, cross-sell, or retention offer. Using a proven combination of real-time predictive analytics and business rules, PredictiveCallCenter automatically provides a recommendation on the agent's screen, along with sales arguments and other information the agent needs to close the sale.

Clementine makes it possible for you to deploy Clementine predictive models directly into SPSS predictive analytic applications like PredictiveCallCenter. This enables you to build sophisticated models with Clementine and deploy them to refine the real-time recommendations delivered by PredictiveCallCenter. Models can be developed to address specific business goals such as improving cross-selling, and even draw upon multi-channel data using predictive analytic capabilities for working with Web, text, and attitudinal data. Multiple predictive models can also be combined, such as models for cross-selling and fraud detection, to ensure that sales efforts aren't focused on customers that are a significant fraud risk.

Deploying models into PredictiveCallCenter

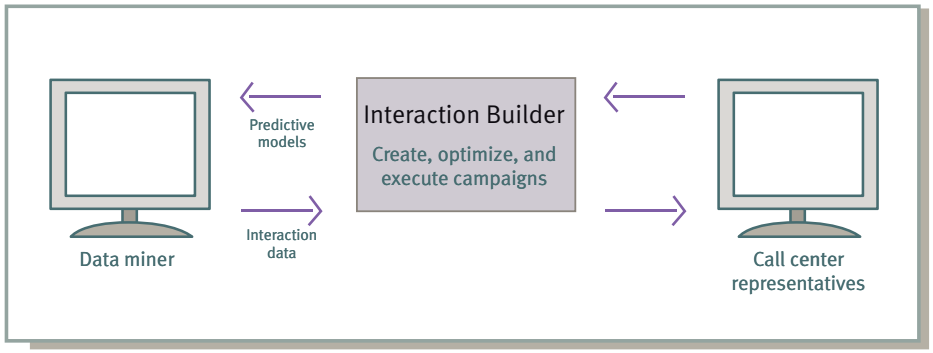


Figure 8: Call center representatives are provided with improved recommendations

Models are deployed into PredictiveCallCenter and other SPSS predictive applications by following a few simple steps in Clementine's predictive applications deployment wizard—immediately providing updated predictive insight to call center representatives (figure 8). Predictive models created using Clementine are deployed into the SPSS Interaction Builder module, an optimization engine that evaluates campaigns at the time of interaction and instantly selects the best recommendation for the customer. SPSS Interaction Builder uses a combination of historical and real-time data entered by the agent during the call itself to choose the best offer. Marketers use SPSS Interaction Builder to build campaigns based on insight from the models. This module enables marketers to determine the optimal timeframe within which to use a model, and the offer-triggering event, such as a customer calling to complain, to associate with the model. New interaction data is continuously made available to Clementine for model updates.

Call center representatives use their existing application—customer data is scored in real-time behind the scenes PredictiveCallCenter's integration with Clementine also enables real-time analysis of text data to improve offers. The notes entered by agents during calls often contain customer complaints, perceptions, and requests that are important predictors of behavior. PredictiveCallCenter "reads" call center notes during each transaction using real-time text mining technology, looking for concepts that may indicate that a customer is at risk of attrition or a good candidate for a cross-sell offer. This information is combined with other customer information to select the best possible offer while the call is in progress.

From service call center to $30 million profit center

Spaarbeleg, a subsidiary of the AEGON Group, the world's third largest life insurance organization, needed to identify sales opportunities in their customer service call center and provide call center agents with the predictive insight needed to increase sales in real time. Spaarbeleg's growth strategy was based on expanding sales to its existing customer base, rather than on acquiring new customers. Aware of the dangers associated with over-saturating its customers with unsolicited messages, Spaarbeleg decided to focus on converting inbound service calls into new sales opportunities. In order for their strategy to succeed, call center agents required targeted guidance to help them identify potential cross-sell opportunities, and support in making the offers.

Originally, the company considered using batch scoring, but this approach would have provided call center agents with outdated recommendations that did not take advantage of new information gathered during the call. Also, batch scoring would not have scaled to call center demands. Spaarbeleg selected PredictiveCallCenter, and integrated it with their existing data warehouse and homegrown call center environment within two months. PredictiveCallCenter transformed Spaarbeleg's inbound call center from a cost center to a profit center—generating an additional $30 million in just one year with a minimal increase in average call time.

Spaarbeleg also uses PredictiveMarketing™ and PredictiveWebSite to optimize other customer interaction channels. PredictiveMarketing increases the response rates of its conventional direct marketing campaigns, and PredictiveWebSite is used to generate leads by placing personalized banners and messages on Spaarbeleg's Web site.

Strategy 4—Customizable real-time

For organizations with decision optimization objectives that require real-time scoring applications usable by different roles within an organization or with the ability address unique business needs, customizable interfaces are critical. Cleo™ and SPSS' Predictive Analytic Framework™ are designed to meet the wide range of business requirements for customizable real-time scoring.

Cleo, a framework for creating Web-based scoring applications, is used to create customized Web applications quickly and easily. To enable multiple users within a company to access Clementine models and score data on demand, companies use Cleo, a Web-based data mining deployment tool. Cleo enables you to cost-effectively deploy Web applications for real-time scoring that put interactive predictive models in the hands of decision makers. Once a predictive model has been created in Clementine, online model deployment is as simple as clicking through the Cleo deployment wizard (figure 9). The Web application is instantly available on the Cleo Server. Decision makers access the Web application to score data in real time whenever they need predictive insight to support their decisions. Unlike many other Web-based analytical tools that require the installation of desktop software or plug-ins, Cleo applications are true thin-client—all that users need for instant access is a Web browser.

Deploying models into Cleo

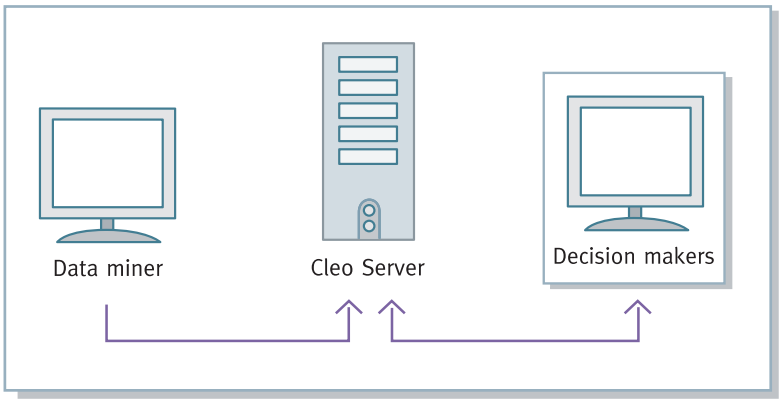


Figure 9: Decision makers access customized Web applications to score data in real time

Organizations that require more advanced functionality often work with SPSS' systems integrator partners to develop more sophisticated Web applications using the Predictive Analytic Framework. This framework is similar to Cleo, in that it is a Web-distributed, thin-client scoring environment that can be used by many different types of users. It also adds additional capabilities such as easy-to-use interfaces for business users to update deployed models or to monitor predictive analytics performance with automated reports such as gains charts (figure 10). As with Cleo, model deployment only requires a few clicks through the Predictive Analytic Framework deployment wizard.

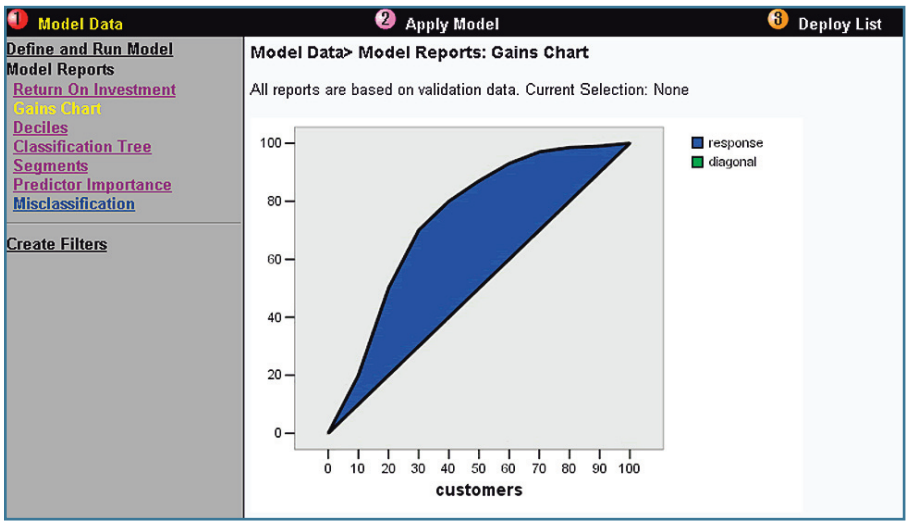


Figure 10: Monitor predictive analytics performance with automated reports



Fighting crime with custom Web applications

Real-time scoring capabilities are often critical to public sector predictive analytics applications such as crime fighting. The Richmond, Virginia, Police Department (PD) uses Clementine data mining to enable their Criminal Analysis Unit to solve serious crimes throughout Virginia's capital city. With Clementine, the Richmond PD accelerates the criminal investigation process to identify minor crimes likely to escalate into violence—helping them send officers where they are needed most.

The police department's criminal analysis group uses the visual and highly intuitive aspects of Clementine to find the patterns and relationships in raw crime data and develop predictive models. These models are then deployed through Cleo over the department's intranet to operational personnel, such as police detectives. The detectives are then able to enter information through a Web interface and receive immediate predictive insight to help focus their crime-fighting efforts.

"Law enforcement and public safety are 24/7 endeavors, and crime frequently occurs when analytical units are not on duty," says Colonel Andre Parker, Richmond PD's Chief of Police. "Web-based analytics, like Cleo, enable law enforcement organizations to deliver analytical capabilities when and where they are needed, while keeping operational personnel on the streets and fighting crime."

Managing predictive models to achieve ROI and compliance goals

The successful execution of decision optimization strategies requires the effective management of deployed models to achieve ROI and compliance goals. As predictive analytics evolves into a mission-critical business process integrated within your IT systems, efficient predictive model management becomes an imperative.

The Clementine data mining workbench provides you with unrivaled capabilities for rapid model development and deployment to create predictive analytics solutions. SPSS Predictive Enterprise Services extends these Clementine capabilities to improve the management of predictive models and related processes within enterprise-wide business operations using a services-oriented architecture. By providing an integrated way to centralize and organize predictive models—and also automate predictive analytics processes—SPSS Predictive Enterprise Services helps you evolve analytical operations beyond ad hoc file systems and inefficient manual updates.

Keep ROI on target by maintaining predictive accuracy

Successfully executing your decision optimization strategy often necessitates the ability to update deployed models and scores easily, as well as control model versions. Premature deployment of models into business operations due to confusion over model status can have a disastrous effect on deployment ROI. Ensure that your deployed predictive models are accurate by tracking and controlling model versions throughout the entire model lifecycle—from development through deployment, maintenance, and retirement.

Models and deployed scores must also be refreshed to achieve your target ROI. The Predictive Enterprise Manager module makes it possible for you to automatically update models with the latest data or re-score databases—either on a regular basis or based on predefined event-trigger conditions. E-mail notifications can be set to alert you when manual intervention is required. All of the SPSS Predictive Enterprise Services automation and version controls help you maintain the predictive accuracy needed to deliver on the target ROI of your predictive analytics initiatives.

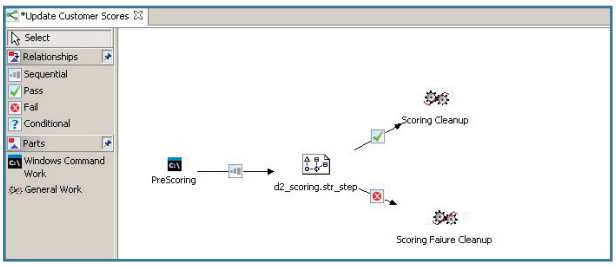


Figure 11: Easily schedule automated model or score updates

Improve compliance with predictive analytics auditing

Predictive models are the center of predictive analytics operations, which often interface with other business-critical processes such as financial reporting and customer relationship management. Recent regulatory requirements, such as those imposed by the Sarbanes-Oxley Act, mandate that effective IT governance controls are in place for auditing these processes. SPSS Predictive Enterprise Services makes predictive analytics auditing possible with complete, detailed model management histories.

Decision optimization strategy next steps

The key to successful data mining is choosing a decision optimization strategy that delivers timely predictive insight wherever it is needed throughout your organization. SPSS predictive analytics experts are available to help you review Clementine's wide range of data mining deployment options and develop a decision optimization strategy that meets your specific business and technical requirements.

Consider the decision optimization strategies and deployment options discussed in this white paper in relation to your business objective:

1. **Rapid-update batch**—Clementine batch mode and Clementine Solution Publisher
2. **Extreme-volume batch**—Custom high-speed scoring engines
3. **Packaged real-time**—SPSS packaged predictive applications such as PredictiveCallCenter and PredictiveWebSite
4. **Customizable real-time**—Customizable Web interface development with Cleo or Predictive Analytic Framework

Contact SPSS Sales for help in choosing the option that will yield the greatest ROI for your organization.

About SPSS Inc.

SPSS Inc. (NASDAQ: SPSS) is the world's leading provider of predictive analytics software and solutions. The company's predictive analytics technology improves business processes by giving organizations consistent control over decisions made every day. By incorporating predictive analytics into their daily operations, organizations become Predictive Enterprises—able to direct and automate decisions to meet business goals and achieve measurable competitive advantage.

More than 250,000 public sector, academic, and commercial customers, including more than 95 percent of the Fortune 1000, rely on SPSS technology to help increase revenue, reduce costs, and detect and prevent fraud. Founded in 1968, SPSS is headquartered in Chicago, Illinois. For additional information, please visit **www.spss.com**.

SPSS

To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners.
© 2005 SPSS Inc. All rights reserved. DMDWP-0705