

➤ Improve Results with High-Performance Data Mining

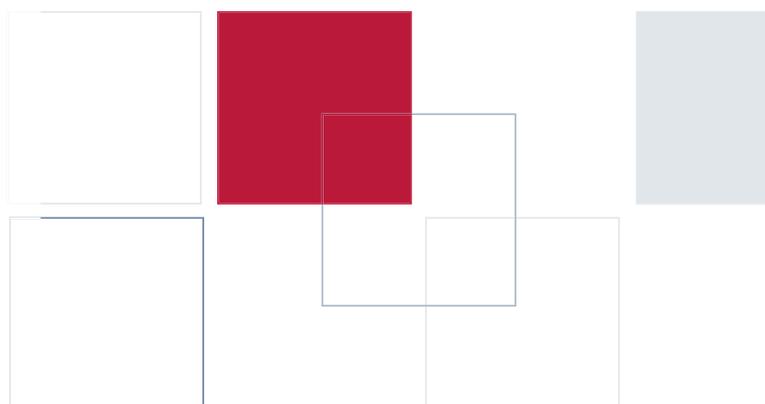
Data mining provides organizations with a clearer view of current conditions and deeper insight into future events. With the Clementine data mining workbench from SPSS Inc., your organization can conduct data mining that incorporates many types of data, resulting in a more detailed view of your customers than other solutions can offer.

Clementine enables your organization to strengthen performance in a number of areas. For example, you can improve customer acquisition and retention, increase customer lifetime value, detect and minimize risk and fraud, reduce cycle time while maintaining quality in product development, and support scientific research.

Use the predictive insight gained through Clementine to guide customer interactions in real time, whether those interactions are taking place in person or through automated systems. You can also manage all of your predictive assets securely and efficiently through Clementine's integration with SPSS Predictive Enterprise Services™. Manage not only your models but also the processes used to arrive at them. This makes data mining more secure and also supports greater analyst productivity.

Clementine is the leading data mining workbench, popular worldwide with data miners and business users alike. By using Clementine, you can:

- Easily access, prepare, and integrate numeric data and also text, Web, and survey data
- Rapidly build and validate models, using the most advanced statistical and machine-learning techniques available
- Efficiently deploy insight and predictive models on a scheduled basis or in real time to the people and systems that make decisions and recommendations



Extend the benefits of data mining throughout your entire enterprise by taking advantage of the integration of Clementine with SPSS Predictive Enterprise Services. SPSS Predictive Enterprise Services enables you to centralize the storage and management of data mining models and all associated processes. You can control the versioning of your predictive models, audit who uses and modifies them, provide full user authentication, automate the process of updating your models, and schedule model execution. As a result, your predictive models become real business assets—and your company gains the highest possible return on your data mining investment.

Clementine offers a number of unique capabilities that make it an ideal choice for today's data-rich organizations.

Streamline the data mining process

Clementine's intuitive graphical interface enables analysts to visualize every step of the data mining process as part of a "stream." By interacting with streams, analysts and business users can collaborate in adding business knowledge to the data mining process. Because data miners can focus on knowledge discovery rather than on technical tasks like writing code, they can pursue "train-of-thought" analysis, explore the data more deeply, and uncover additional hidden relationships.

From this visual interface, you can easily access and integrate data from textual sources, data from Web logs, and—with Clementine 10.0—data from Dimensions™ survey research products. No other data mining solution offers this versatility.

Choose from an unparalleled breadth of techniques

Clementine offers a broad range of data mining techniques that are designed to meet the needs of every data mining application, producing the world's most popular data mining workbench. You can choose from a number of algorithms for clustering, classification, association, and prediction. These algorithms can be regularly calibrated and verified to ensure that they support the creation of powerful data mining models.

Optimize your current information technologies

Clementine is an open, standards-based solution. It integrates with your organization's existing information systems, both when accessing data and when deploying results. You don't need to move data into and out of a proprietary format. This helps you conserve staff and network resources and deliver results faster.

With Clementine, you can access data in virtually any type of database, spreadsheet, or flat file. In addition, Clementine performs in-database modeling and scoring in major databases offered by IBM®, Oracle®, and Microsoft®, which improves the speed and scalability of these operations. Clementine can deploy results to any of SPSS' predictive applications, as well as to solutions provided by other companies. This interoperability helps your IT department meet internal customer needs while ensuring that your organization gains even greater value from your information technology investments.

Leverage all your data for improved models

Only with Clementine can you directly and easily access text, Web, and survey data, and integrate these additional types of data in your predictive models. SPSS customers have found that using additional types of data increases the “lift” or accuracy of predictive models, leading to more useful recommendations and improved outcomes.

With the fully integrated Text Mining for Clementine® module, you can extract concepts and opinions from any type of text—such as internal reports, call center notes, customer e-mails, media or journal articles, blogs, and more. And, with Web Mining for Clementine®, you can discover patterns in the behavior of visitors to your Web site.

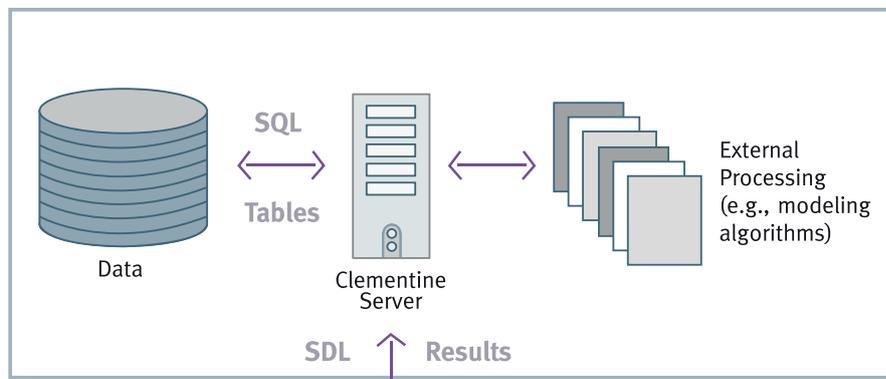
Direct access to survey data in Dimensions products enables you to include demographic, attitudinal, and behavioral information in your models—rounding out your understanding of the people or organizations you serve.

Deploy data mining results efficiently

Clementine is scalable—capable of cost effectively analyzing as few as several thousand records or as many as tens of millions. It uses a three-tiered architecture to manage modeling and scoring with maximum efficiency.

Clementine’s deployment options enhance its scalability by enabling your organization to deliver results to suit your particular requirements. For example, through Clementine’s batch mode feature or the Clementine Solution Publisher™ component, you can deploy data mining streams to your database to run in the background of other operational processes. Scores or recommendations are created without interrupting daily business tasks.

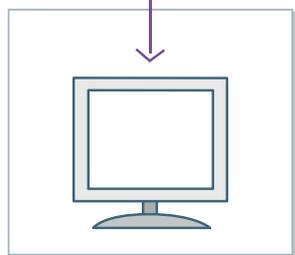
■ Server



Clementine provides data mining scalability by using a three-tiered architecture, as shown in this diagram. The Clementine Client tier (shown at the bottom) passes stream description language (SDL) to Clementine Server. Clementine Server then analyzes particular tasks to determine which it can execute in the database.

After the database runs the tasks that it can process, it passes only the relevant aggregated tables to Clementine Server.

■ Client

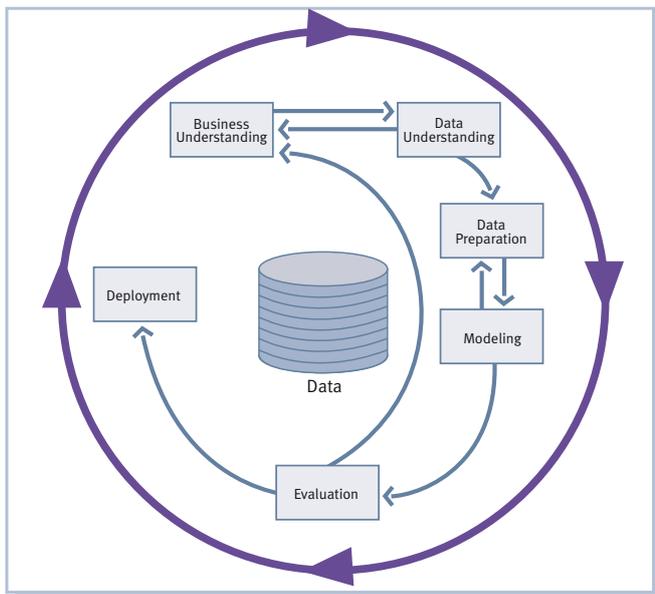


Or you may choose to export Clementine models to other applications in industry-standard Predictive Model Markup Language (PMML). You can also export information about the steps in the process, such as data access, modeling, and post-processing. This option gives you greater flexibility in how you process models and integrate predictive insight with other systems.

Through its integration with all of SPSS' predictive applications and its openness to other systems, Clementine supports the delivery of recommendations to managers, customer-contact staff, and the systems supporting your operations. More information about SPSS' predictive applications, including PredictiveCallCenter™, PredictiveClaims™, PredictiveMarketing™, and PredictiveWebSite™, can be found at www.spss.com/predictive_analytics.

Follow a proven, repeatable process

During every phase of the data mining process, Clementine supports the de facto industry standard, the Cross-Industry Standard Process for Data Mining (CRISP-DM). This means your company can focus on solving business problems through data mining, rather than on reinventing a new process for every project. Individual Clementine projects can be efficiently organized using the CRISP-DM project manager. Or, with SPSS Predictive Enterprise Services, you can support the CRISP-DM process enterprise-wide.



The CRISP-DM process, as shown in this diagram, enables data miners to implement efficient data mining projects that yield measurable business results.

Major enhancements in Clementine 10.0

With this release, SPSS continues its commitment to delivering a data mining solution that offers the greatest possible efficiency and flexibility in the development and deployment of predictive models.

Enhancements available in Clementine 10.0 offer your organization:

- Greater productivity
 - A new anomaly detection algorithm enables analysts to detect unusual cases faster than they could before
 - New feature selection capabilities automatically identify the most- and least-relevant fields for a particular analysis
 - Multiple enhancements have been made to make data preparation easier and more efficient
 - Improvements to the Clementine interface make it even more intuitive and easy to use
- Increased performance
 - Clementine 10.0 supports in-database caching, database write-back with indexing, and optimized merging for joining tables outside of the database
 - It also supports the use of parallel processing during a number of data pre-processing and model building operations for organizations with multiple CPUs or multi-core CPUs in their operating environment

- Additional integration
 - Clementine 10.0 provides direct access to data collected using SPSS' Dimensions family of survey research products
 - It now supports the Red Hat® Linux operating system, in addition to numerous other platforms
 - It offers additional integration with Excel® and other Microsoft Office products

This release further enhances Clementine's integration with SPSS Predictive Enterprise Services, a solution that enables your organization to centralize analytic assets in a secure, auditable repository for more efficient management and more effective use of predictive models enterprise-wide.

Now, for stronger support of the CRISP-DM process, you can automatically save entire projects, including all process steps, to SPSS Predictive Enterprise Services, and update models from within the repository interface.

Features

Clementine's main features are described below in terms of the CRISP-DM process.

Business understanding

Clementine's visual interface makes it easy for your organization to apply business knowledge to data mining projects. In addition, optional business-specific Clementine Application Templates (CATs) are available to help you get results faster. CATs ship with sample data that can be installed as flat files or as tables in a relational database schema.

- CRM CAT*
- Telco CAT*
- Fraud CAT*
- Microarray CAT*
- Web Mining CAT* (requires the purchase of Web Mining for Clementine)

Data understanding

- Obtain a comprehensive first look at your data using Clementine's data audit node
- View data quickly through graphs, summary statistics, or an assessment of data quality
- Create basic graph types, such as histograms, distributions, line plots, and point plots
- Employ additional graph types, such as box plots, heat maps, scatterplot matrices, linkage analysis plots, and more, through Advanced Visualization for Clementine
- Use association detection when analyzing Web data
- Interact with data by selecting a region of a graph and see the selected information in a table; or use the information in a later phase of your analysis

Data preparation

- Access data
 - Structured (tabular) data
 - Access ODBC-compliant data sources with the included SPSS Data Access Pack. Drivers in this middleware pack support IBM DB2®, Oracle, Microsoft SQL Server™, Informix®, and Sybase® databases.
 - Import delimited and fixed-width text files, any SPSS® file, and SAS® 6, 7, 8, and 9 files
 - Specify worksheets and data ranges when accessing data in Excel

- Unstructured (textual) data
 - Automatically extract concepts from any type of text by using Text Mining for Clementine*
- Web site data
 - Automatically extract Web site events from Web logs using Web Mining for Clementine*
- Survey data
 - Directly access data stored in the Dimensions Data Model or in the data files of Dimensions* products
- Data output
 - Work with delimited and fixed-width text files, ODBC, Microsoft Excel, SPSS, and SAS 6, 7, 8, and 9 files
 - Export in XLS format through the Excel Output Node
- Choose from various data-cleaning options
 - Remove or replace invalid data
 - Automatically fill in missing values
- Manipulate data
 - Work with complete record and field operations, including:
 - Field filtering, naming, derivation, binning, re-categorization, value replacement, and field reordering
 - Record selection, sampling, merging (through inner joins, full outer joins, partial outer joins, and anti-joins), and concatenation; sorting, aggregation, and balancing; deriving new fields based on conditional criteria; and calculating new fields
 - Specialized manipulations for showing the “history” of values and converting set variables into flag variables
 - Filling a generated field with the value of another field using the Restructure node
 - Converting fields to records with the Transpose node
 - Binning records by equal sum of the binned variable, rather than by equal record count
 - Using new string functions: string creation, substitution, search and matching, whitespace removal, and truncation
 - Preparing data for time-series analysis with the Time Plot node
- Partition data into training, test, and validation datasets

Modeling

- Mine data in the database where it resides, with in-database modeling. Support:
 - IBM DB2 Enterprise Edition 8.2 decision trees, regression, association, and demographic clustering techniques
 - Oracle 10g Naïve Bayes and Adaptive Bayes networks and Support Vector Machines (SVM)
 - Microsoft SQL Server Analysis Services decision trees
- Use predictive and classification techniques
 - Neural networks (multi-layer perceptrons using error back-propagation, radial basis function, and Kohonen networks)
 - Browse the importance of the predictors
 - Decision trees and rule induction techniques, including CHAID, exhaustive CHAID, QUEST, and C&RT
 - Browse and create splits in decision trees interactively
 - Rule induction techniques in C5.0
 - Browse, collapse, and expand decision rules
 - Linear regression, logistic regression, and multinomial logistic regression
 - View model equations and advanced statistical output
 - Anomaly Detection node for directly detecting unusual cases
 - Feature Selection node for identifying fields that are most and least important to an analysis
 - Additional statistics available at the Means and Matrix nodes
- Use clustering and segmentation techniques
 - Kohonen networks, K-means, and TwoStep
 - View cluster characteristics with a graphical viewer
- Choose from several association detection algorithms
 - GRI, Apriori, sequence, and CARMA algorithms
 - Score data using models generated by association detection algorithms
 - Filter, sort, and create subsets of association models using the association model viewer

Features subject to change based on final product release. □ Symbol indicates a new feature. * Separately priced modules

- Employ data reduction techniques
 - Factor analysis and principal components analysis
 - View model equation and advanced statistical output
- Combine models through meta-modeling
 - Multiple models can be combined, or one model can be used to build a second model
- Import PMML-generated models created in other tools such as AnswerTree® and SPSS for Windows®
- Use Clementine External Module Interfaces (CEMI) for custom algorithms
 - Purchase add-on tools from the Clementine Partner Plus Program

Evaluation

- Easily evaluate models using lift, gains, profit, and response graphs
 - Use a one-step process that shortens project time when evaluating multiple models
 - Define hit conditions and scoring expressions to interpret model performance
- Analyze overall model accuracy with coincidence matrices and other automatic evaluation tools

Deployment

Clementine offers a choice of deployment capabilities to meet your organization's needs.

- Clementine Solution Publisher (optional*)
 - Automate the export of all operations, including data access, data manipulation, text mining, model scoring—including combinations of models—and post-processing
 - Use a runtime environment for executing image files on target platforms
- PredictiveCallCenter, PredictiveClaims, PredictiveMarketing, or PredictiveWebSite™ (optional*)
 - Automatically export Clementine streams
 - Combine exported Clementine streams with predictive models, business rules, and exclusions to optimize customer interactions

- Cleo (optional*)
 - Implement a Web-based solution for rapid model deployment
 - Enable multiple users to simultaneously access and immediately score single records, multiple records, or an entire database, through a customizable browser-based interface
- Clementine Batch
 - Automate production tasks while working outside the user interface
 - Automate Clementine processes from other applications or scheduling systems
 - Generate encoded passwords
 - Call Clementine processes via the command line
- Scripting
 - Automate command-line scripts or scripts associated with Clementine streams to automate repetitive tasks in the user interface. Scripts generally perform the same types of actions that you otherwise would carry out using a mouse or keyboard.
 - Execute selected lines from a stream, SuperNode, or stand-alone script using an icon on the toolbar
 - Update stream parameters within a stand-alone script
- Export generated models as PMML 2.1
 - Perform in-database scoring, which eliminates the need for—and costs associated with—transferring data to client machines or performing calculations there
 - Deploy Clementine PMML models to IBM DB2 Intelligent Miner™ Visualization and Intelligent Miner Scoring
- Use the bulk-loading capabilities of your database
 - Increase performance during data export by using your database's bulk loader. Fine-tune various options, including row-wise or column-wise binding for loading via ODBC, and batch-size settings for batch commits to the database.

SPSS Predictive Enterprise Services** (optional*)

- Centralize data mining projects to leverage organizational knowledge
 - Save streams, models, and other objects in a central, searchable repository
 - ▢ Save entire Clementine projects, including all process steps
 - Group streams in folders, and secure folders and streams by user or user groups
 - Provide permission-based access to protect the privacy of sensitive information
 - Reuse the most effective streams and models to improve processes and increase the accuracy of results
 - Search on input variables, target variables, model types, notes, keywords, authors, and other types of metadata
 - ▢ Automate Clementine model refreshment from within SPSS Predictive Enterprise Services
- Ensure reliable results by controlling versions of predictive models
 - Automatically assign versions to streams and other objects
 - Protect streams from being overwritten through automatic versioning

Scalability

- Use in-database mining to leverage parallel database implementations
- Use in-database modeling to build models in the database using leading database technologies
- ▢ Support multi-threaded processing during sorting and model building, in environments with multiple CPUs or multi-core CPUs
- Minimize network traffic via intelligent field projection, which means that Clementine pulls data only as needed from your data warehouse and passes only relevant results to the client

** SPSS Predictive Enterprise Services was previously referred to as SPSS Model Manager

For more than 25 years, U.S.-based National Instruments has revolutionized the way engineers and scientists in industry, government, and academia approach measurement and automation. The company sells virtual instrumentation software and hardware in a wide range of industries, from automotive to consumer electronics.

As National Instruments began to rapidly expand from its mid-sized roots, it recognized a need for more sophisticated data mining software to keep up with accompanying growth in sales leads and customers. Specifically, the company sought to identify the best prospects for its salespeople to target and segment customers for more strategic marketing programs.

“Our customer base was so broad and diverse across geography, industry, and even from customer to customer, that we didn’t know if it could be segmented. By using Clementine, we now have a robust profile of each of our market segments, which allows us to use direct marketing to customize messages delivered to each one.”

– *Amanda Jensen*
Analytics Manager
Business Intelligence Department
National Instruments Corporation

The cornerstone of the Predictive Enterprise

Clementine makes data predictive and facilitates the delivery of predictive insight to the people in your organization who make decisions and the systems that support daily customer interactions.

If your organization has any data stored in text files or in Web logs—as many do—you can leverage it by using Text Mining for Clementine or Web Mining for Clementine. And you can understand the attitudes and beliefs that lie behind behavior—why customers make the choices they do—by incorporating survey data from any of SPSS’ Dimensions survey research products.

Thanks to its integration with SPSS predictive applications and other information systems, Clementine enables you to guide daily decisions and recommendations, as well as long-range planning, with insight into current and future conditions. You can accomplish this securely and efficiently, across your entire enterprise, with SPSS Predictive Enterprise Services.

Clementine's extensive capabilities are supported by the most advanced statistical and machine learning techniques available. For greater value, it is open to operation and integration with your current information systems.

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. CLM10SPCA4-1005

