

➤ Achieve Better Insight and Prediction with Data Mining

Data mining provides organizations with a clearer view of current conditions and deeper insight into future events. With Clementine from SPSS Inc., your organization can conduct data mining that incorporates many types of data, resulting in more in-depth knowledge of every aspect of your operations—including more complete analysis and understanding of your customers and constituents.

Clementine enables your organization to strengthen performance in a number of areas. For example, you could:

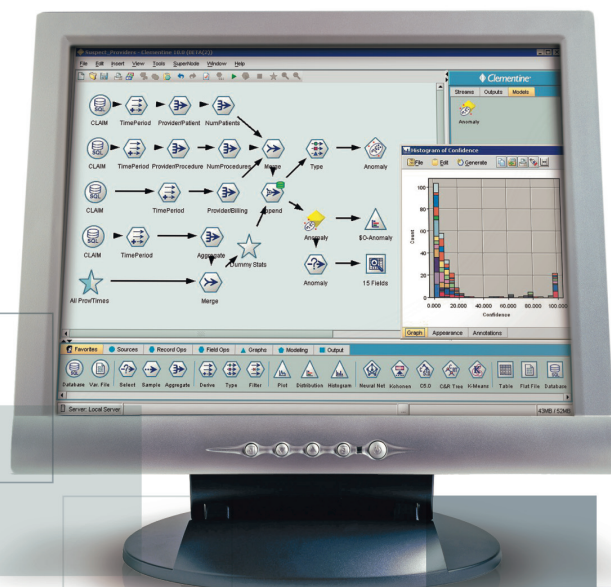
- Improve customer acquisition and retention
- Increase customer lifetime value
- Detect and minimize risk and fraud
- Reduce cycle time while maintaining quality in product development
- Support scientific research

Use the predictive insight gained through Clementine to guide customer interactions in real time and share that insight throughout your organization. With Clementine's powerful data preparation, visualization, and predictive modeling capabilities, you can solve business problems faster.

Clementine is popular worldwide with data miners and business users alike because it enables you to:

- Easily access, prepare, and integrate structured 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 that make decisions and recommendations, and the systems that support them

Clementine's many unique capabilities make it an ideal choice for addressing the business applications found in 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, Web logs, SPSS’ Dimensions™ survey research products, as well as data in virtually any type of database, spreadsheet, or flat file—including SPSS, SAS®, and Microsoft® Excel® files. No other data mining solution offers this versatility.

Clementine’s powerful automation tools make it easy for data miners to find the best model, based on hidden patterns in their data, and quickly produce consistent and accurate results.

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.

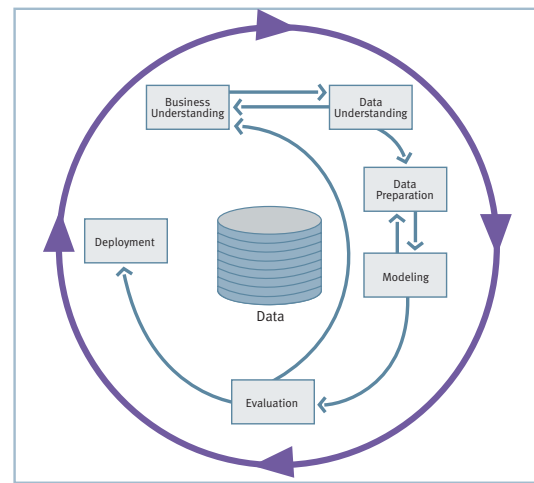
Direct access to survey data in Dimensions products enables you to include demographic, attitudinal, and behavioral information in your models—rounding out your understanding of the people or organizations you serve.

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. You can choose from a number of algorithms for prediction, clustering, and association, including survival analysis, neural networks, support vector machines, and graphical probabilistic modeling.

Optimize your current information technologies

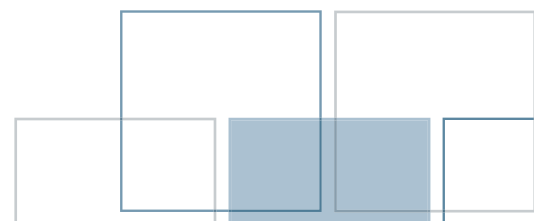
Clementine is an open, standards-based solution. It integrates with your organization’s existing information systems, both when accessing data and when deploying results. You don’t need to move data into and out of a proprietary format. This helps you conserve resources, deliver results faster, and reduce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



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

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.



Add enterprise-level capabilities

Clementine can efficiently analyze the amounts of data typically generated by small to mid-sized organizations. If your data mining needs grow in volume or complexity, SPSS makes it easy for you to move to our enterprise-level offering, Clementine Server.

Using a client/server architecture, Clementine Server enables multiple data analysts to work simultaneously without straining computing resources. You can take advantage of in-database data mining on leading information platforms and efficiently process large amounts of data. Clementine Server also offers additional deployment options—helping you to extend the benefits of data mining across geographic or functional lines and put results in the hands of decision makers.

You can further optimize analytical assets throughout your organization by using Clementine with SPSS Predictive Enterprise Services™, which enables you to centralize the storage and management of data mining models and all associated processes. With this platform, 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 organization gains the highest possible return on your data mining investment.

What's new in Clementine 12.0

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.

Clementine features extensive enhancements in several key areas to help your organization improve performance, productivity, and return on your data mining investment.

Unified customer analysis techniques

Gain a complete customer analysis toolkit that unifies both advanced and traditional techniques in a single framework.

- Identify your best customers using the RFM customer value segmentation scoring technique
- Accurately estimate customer attrition using Cox regression for survival analysis
- Easily combine these traditional techniques with advanced data mining prediction and segmentation algorithms

Automated modeling

Enhance productivity and achieve faster time-to-solution using Clementine's powerful automated modeling.

- Find the best models for predicting both binary (“yes/no”) and numeric outcomes using automated modeling operations that create and evaluate many different models in a single step
- Select the best model more easily with visual evaluation information
- Gain more detailed control over automated modeling using new frequency, weights, and misclassification costs features

Enhanced analytics

Leverage new analytical techniques to improve predictive accuracy and achieve better results.

- Obtain more accurate predictions by combining two or more models with the new Ensemble node
- Employ advanced sampling techniques, including stratified sampling to obtain more representative results, and clustered sampling to ensure samples include all related items
- Compare scores easily across different algorithms and technologies using propensity scores

Wider range of algorithms

Solve more business problems and predict outcomes with greater accuracy using several new algorithms.

- Make more accurate predictions when working with wide data sets using the new Support Vector Machine (SVM) node
- Obtain richer insight from graphical Bayesian network models

Improved visualization and reporting

Use Clementine's new reporting and visualization capabilities to achieve better insight and more effective communication. Produce the right graph or table type, export analysis to third-party tools, and communicate results clearly across the organization.

- Create more compelling graphs using a wizard-like interface that guides you to the most appropriate chart types for your data
- Interact with graphs via rich data selection tools, resulting in more accurately focused analysis
- Improve your reporting by including easy-to-create custom tables. Nest, stack, or layer variables in multiple dimensions to display summaries for multiple statistics and display multiple response sets.
- Discover the best models and get better predictive insight with variable importance charts, which rank variables according to their relative importance

Improved scalability and integration

With Clementine 12.0, you gain both desktop tools and an upgrade path to a highly scalable, highly integrated, and easily managed data mining and predictive analytics platform for the enterprise.

- Leverage extended database optimization for improved performance
- Gain enhanced real-time scoring capabilities
- Achieve a deeper level of integration with custom or third-party components

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 a great deal of data stored in text form or in Web logs, you can gain additional value from this data 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.



“We are extremely impressed with the customer-centric response from SPSS in advancing their data mining product to help organizations, such as EarthLink, better meet their customer marketing needs.”

– *Atique Shah*
Vice President of Direct Marketing
and Consumer Insights
EarthLink, Inc.



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 so that you can easily see the details of best-practice techniques.

- 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
- ▣ Create a wide range of basic and advanced graphs with automatic assistance using the Graphboard node
- ▣ Create complex tabular reports easily using the Custom Table node
- Edit your graphs to communicate results more clearly
- Use visual link analysis to see the associations in your data
- ▣ Interact with data by selecting regions or items on a graph and view the selected information; or use it in a later stage of your analysis
- Access SPSS statistics, graphs, and reporting tools directly from Clementine

Data preparation

- Access a wide range of data
 - Structured (tabular) data
 - ODBC-compliant data sources with the 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 use Dimensions Data Model™ or in the data files of Dimensions* products
- Data export
 - Export data to delimited text files, Microsoft Excel, SPSS, SAS 6, 7, 8, and 9 files, and many databases
 - Export in XLS format through the Excel Output Node
 - ▣ Export data to Dimensions* for survey applications
- Choose from various data-cleaning options
 - Remove or replace invalid data
 - Use predictive modeling to automatically impute missing values
 - Automatically generate operations for the detection and treatment of outliers and extremes
- 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 (including clustered and stratified sampling), merging (including inner joins, full outer joins, partial outer joins, and anti-joins), and concatenation; sorting, aggregation, and balancing
 - Data restructuring, including transposition
 - ▣ Binning numerical attributes into sub-ranges optimized for prediction
 - Extensive string functions: string creation, substitution, search and matching, whitespace removal, and truncation
 - Preparing data for time-series analysis
- Partition data into training, test, and validation datasets
- Transform data automatically for multiple variables
 - Visualization of standard transformations

- Access data management and transformations performed in SPSS directly from Clementine
- ▣ RFM scoring: aggregate customer

transactions to provide Recency, Frequency, and Monetary value scores and combine these to produce a complete RFM analysis.

Modeling

Employ a wide range of data mining algorithms with many advanced features to get the best possible results from your data.

- Use interactive model and equation browsers and view advanced statistical output
- ▣ Show relative impact of different data attributes on predicted outcomes with variable importance graphs
- Combine models through meta-modeling
 - Multiple models can be combined, or one model can be used to analyze a second model
 - ▣ Ensemble node combines predictions automatically for improved accuracy
- Import PMML models from other tools such as AnswerTree® and SPSS
- ▣ Use the Clementine Extension Framework (CLEF)* for custom algorithms

Clementine's data mining algorithms are organized into a "base" module and optional additional algorithm modules. The base module includes:

- C&RT, CHAID & QUEST—Decision tree algorithms including interactive tree building
- Decision List—Interactive rule-building algorithm enables you to incorporate business knowledge into a predictive model
- K-means—Clustering
- GRI—Generalized rule induction association discovery algorithm
- Factor/PCA—Data reduction using factor analysis and principal component analysis
- Linear Regression—Best-fit linear equation modeling

The Clementine Classification Module* includes:

- ▣ Binary classifier and numeric predictor—Automate the creation and evaluation of multiple models
- Self-learning response model—Bayesian model with incremental learning
- Time-series—Generate and automatically select time-series forecasting models

- C5.0 decision tree and rule set algorithm
- Neural Networks—Multi-layer perceptrons with back-propagation learning, and radial basis function networks
- ▣ Support Vector Machines—Advanced algorithm with accurate performance for wide datasets
- ▣ Bayesian Networks—graphical probabilistic models
- ▣ Cox regression—calculate likely time to an event
- Binomial and multinomial logistic regression
- Discriminant analysis
- General linear models (GLM)

The Clementine Segmentation Module* includes:

- Kohonen Network—Clustering neural network
- TwoStep Clustering—Select the right number of clusters automatically
- Anomaly Detection—Detect unusual records through the use of a cluster-based algorithm

The Clementine Association Module* includes:

- Apriori—Popular association discovery algorithm with advanced evaluation functions
- CARMA—Association algorithm which supports multiple consequents
- Sequence—Sequential association algorithm for order-sensitive analyses

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
- ▣ Use propensity scores for consistent deployment and easy comparison between diverse model types

Deployment

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

- Export models using SQL or PMML (the XML-based standard format for predictive models)
- Clementine Solution Publisher Runtime™ (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
- Automatically export Clementine streams to SPSS predictive analytics applications
 - Combine exported Clementine streams and predictive models with 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
- Scripting
 - Use scripts to automate complex repetitive tasks

Clementine Server (optional*)

Clementine Server, SPSS' enterprise-level offering, provides all of the data mining capabilities of the client version plus increased performance and other functionality. For a complete list of features, refer to the Clementine Server specifications sheet. Key capabilities enable you to:

- Employ in-database mining to leverage high-performance database implementations
- Use in-database modeling to build models in the database using leading database technologies (IBM® DB2® Enterprise Edition 8.2, Oracle 10g, and Microsoft SQL Server

Analysis Services) and leverage high-performance database implementations

- Vendor-supplied algorithms included for IBM DB2 DWE, Oracle Data Mining, and Microsoft SQL Server 2005

- Leverage high-performance hardware, experience quicker time-to solution, and achieve greater ROI through parallel execution of streams and multiple models
- Transmit sensitive data securely between Clementine Client and Clementine Server through secure sockets layer (SSL) encryption

SPSS Predictive Enterprise Services (optional*)

SPSS Predictive Enterprise Services is an enterprise-level platform that enables you to manage and automate your analytical processes and easily deploy results across your organization to increase productivity and increase the value of your analytical investment. SPSS Predictive Enterprise Base Services helps you to:

- Centralize and manage analytical assets to leverage organizational knowledge and provide powerful change management capabilities to help with auditability and compliance
- Automate your analytical processes to increase productivity and ensure reliable, consistent, accurate results
- Deploy analytical results by delivering output through customizable, browser-based end-user interfaces or integrating with your existing applications using standardized Web services interfaces

SPSS Predictive Enterprise Services Clementine Client Adapter enables analysts to interact directly with the Base services to store, retrieve, browse, and search for analytical assets. SPSS Predictive Enterprise Services Clementine Server Adapter enables the Process Manager to control Clementine tasks. For more information about how your organization can benefit by using Clementine Server with SPSS Predictive Enterprise Services, see the SPSS Predictive Enterprise Services brochure.

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

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. © 2008 SPSS Inc. All rights reserved. CLM125PC-0108

