IBM SPSS Modeler Premium

Improve model accuracy with unstructured data

IBM® SPSS® Modeler Premium helps you achieve superior outcomes by basing your business decisions on predictive intelligence. Predictive intelligence creates more effective strategies because it allows organizations not only to evaluate trends and benchmarking results, plans and performance, but also to look into the future by evaluating likely outcomes and understanding how the interplay of factors affects those outcomes.

With SPSS Modeler Premium, you can solve any business problem faster using powerful, proven analytical techniques that deliver deeper insight into your customers or constituents. However, because the majority of data is trapped in unstructured or textual form—comments, files or on the web—modeling with structured data alone may provide an incomplete view into your business processes and outcomes.

SPSS Modeler Premium enables organizations to tap into the predictive intelligence held in all forms of data. It goes beyond the analysis of structured numerical data and includes information from unstructured data, such as web activity, blog content, customer feedback, emails, articles and more to discover relationships between concepts and sentiments and create the most accurate predictive models possible.

Boost productivity, quality and accuracy

SPSS Modeler Premium allows you to perform both text analysis and data mining within an interactive, visualization-based environment. The intuitive graphical interface makes it easy to see every step of the data mining process as part of a “stream.” Text analytics is straightforward and efficient, with interactive graphs to help you explore and display text data and patterns for instant analysis, as well as powerful classification and categorization techniques that transform text into an analytical asset.

Highlights

Solve business problems faster with analytical techniques that deliver deeper insight.

- Easily access, prepare and integrate structured data and text, web and survey data.
- Quickly identify and extract sentiments from text in more than 30 languages.
- Extend access to text stored in legacy systems, such as mainframes, through support for IBM Classic Federation Server and IBM® zDB2® data stores.
- Easily perform in-database analytics and mining for IBM® Netezza®.
- Broaden visibility into the full customer view with the ability to export to IBM® Cognos® Business Intelligence directly from the interface.
- Add more deployment options through zLinux, SuSE Linux Enterprise Server and inclusion in IBM Smart Analytics System for Power.
Enhancements available in the latest release make it easier to extend the reach of text analytics throughout your organization. Read more about these enhancements on page two.

From the SPSS Modeler Premium visual interface, you can easily access and integrate data from many sources, including data in virtually any type of database, spreadsheet or flat file (such as IBM SPSS Statistics, SAS and Microsoft Excel files), textual data and data from Web 2.0 sources (such as RSS feeds), IBM SPSS Data Collection products files, and Cognos Business Intelligence and legacy systems with IBM Classic Federation Server and zDB2 support. No other data mining solution offers this versatility.

In addition, SPSS Modeler Premium offers in-database mining for IBM InfoSphere, Microsoft SQL Server, Oracle and, with the latest release, IBM Netezza. With in-database mining, there is no need to move data from large databases to SPSS Modeler Premium Server; the analytics and mining take place in-database. This results in a significant improvement in analytical performance.

With powerful automation tools such as automated data preparation and auto modeling, SPSS Modeler Premium makes it easy to prepare data for analysis, find the best model based on hidden patterns in the data and quickly produce consistent and accurate results. Industry-specific text analysis packages and templates speed the analysis process and ensure you obtain the most accurate results. Powerful natural language processing capabilities help structure text into hierarchical categories that can be integrated into predictive models automatically.

The solution supports the CRoss-Industry Standard Process for Data Mining (CRISP-DM), which enables analysts to focus on solving business problems, rather than on programming. Individual projects can be efficiently organized using the CRISP-DM project manager.

### What’s new in IBM SPSS Modeler Premium

The enhancements to the latest version of SPSS Modeler Premium bring predictive intelligence to a broader range of organizations, enabling them to be more focused and agile in their planning and daily decision making because they will have a more nuanced understanding of their own enterprises, the environment in which they operate and of their customers and other stakeholders. New capabilities include:

- **IBM Cognos integration**. Analysts can access structured data directly through Cognos Business Intelligence software. IBM Cognos software organizes and delivers a complete and consistent view of information for enterprise-wide decision making. By adding the analytic capabilities of SPSS Modeler Premium, organizations can quickly and reliably evaluate the likelihood of specific outcomes. Also – since SPSS Modeler Premium can write results into Cognos Business Intelligence – they can make predictive intelligence gleaned from a customer view that combines structured and unstructured data available to business users and all information stakeholders who rely on Cognos dashboards.

- **IBM Netezza functionality**. The latest version of SPSS Modeler Premium introduces the capability to perform in-database analytics and mining within the Netezza appliance. Netezza advanced analytics technology fuses data warehousing and analytics into a high-performance appliance that enables organizations to build and deploy analytic applications that scale to sizes only addressable by supercomputers today.

- **Enhanced InfoSphere™ Warehouse and IBM DB2® functionality**. Organizations with an InfoSphere Data Warehouse can use the full range of its data mining algorithms from within SPSS Modeler. This gives users the advantage of an intuitive, graphical interface without compromising data mining performance. Supported algorithms now include logistic regression, Naïve Bayes, time-series and radial basis function (RBF).
• **Mining against mainframe data**. Mainframe computers typically hold a wealth of data about an organization’s operational history. SPSS Modeler can directly access data on systems using the IBM System z® operating system, allowing organizations to see what effect past events will have on the future. This will help them understand current activity more clearly, understand changing business realities and base their planning on reliable predictive intelligence.

• **Support for Linux on System z**. Running Linux on System z offers many advantages to organizations seeking to simplify complex information systems while meeting today’s requirements for security, transparency and cost control. SPSS Modeler Premium now is supported in these environments.

In addition, enhancements added in SPSS Modeler Premium 14 strengthened its ability to extract insights from text data, enabling organizations to:

- Receive faster, more accurate results when analyzing banking, insurance or advertising text, emoticons and slang with new industry-specific text analysis packages and templates.
- Create hierarchical categorization structures to organize concepts more logically and in greater detail.
- Import pre-defined categories, including hierarchical categories, annotations and keyword descriptors, and export them to Excel.
- Save hierarchical categories for reuse with an enhanced semantic network grouping technique for category building.
- Extract text faster and more accurately, especially when working with large datasets, by leveraging new industry-sensitive semantic networks and more efficient use of hardware.
- Define and test rules on sample text before applying them to your data, using the enhanced text link rule editor.

**Explore a broader range of data**

Only with SPSS Modeler Premium can you directly access text, web and survey data and integrate these additional types of data into your predictive models for more useful recommendations and improved outcomes.

The interactive text mining workbench will save you time and effort while helping you gain a competitive advantage. Unlike other text analytics tools, you do not need a linguistic background to use it. You can easily customize concept dictionaries for a particular domain area by using the Resource Editor, an integrated resource for managing the text extraction process. This enables you to find relevant concepts and associations faster.

You can also create customized templates and libraries for specific business applications directly from the main SPSS Modeler Premium toolbar, and reuse these valuable resources with other products and applications, including IBM SPSS Text Analytics for Surveys. The linguistic resources in SPSS Modeler Premium support a range of industries and applications, including sentiment analysis, CRM, security and intelligence, market intelligence, life sciences (genomics and MESH) and IT.

**Choose from an unparalleled breadth of techniques**

SPSS Modeler offers an array of advanced data mining techniques that are designed to meet the needs of every data mining application, including the following data mining algorithms:

- **Classification algorithms**. Make predictions or forecasts based on historical data using techniques such as Decision Tree, Neural Networks, Logistic Regression, Time-Series, Support Vector Machines, Cox regression and more. Leverage automatic classification modeling for both binary and numeric outcomes to streamline model creation.

- **Segmentation algorithms**. Group people or detect unusual patterns with automatic clustering, anomaly detection and clustering neural network techniques. Use automatic classification to apply multiple algorithms with a single step and take the guesswork out of selecting the right technique.

- **Association algorithms**. Discover associations, links or sequences using apriori, CARMA and sequential association.
Through a proven natural language processing (NLP) linguistic extraction process, SPSS Modeler Premium pulls key concepts from many types of unstructured data and groups them into categories. Extracted concepts, opinions and categories are then combined with structured data and applied to predictive models to uncover valuable insights into actions, behaviors, patterns and associations.

Text link analysis (TLA) technology helps you identify and extract sentiments and opinions in multiple languages, including Dutch, English, French, German, Italian, Japanese, Portuguese and Spanish text. The English-language version also includes an interface that supports third-party translation options.

**Optimize your current information technologies**

The open and scalable architecture of SPSS Modeler Premium makes the best use of your existing IT infrastructure. It integrates with your existing systems, both when accessing data and when deploying results, so you don’t need to move data into and out of a proprietary format. And techniques such as in-database modeling, multithreading, clustering and embedded algorithms help you conserve resources, deliver results faster and reduce overall IT costs.

**Support data mining across the enterprise**

SPSS Modeler Premium can efficiently analyze the amounts of data typically generated by small to mid-sized organizations. Organizations with high-volume or complex data mining requirements can leverage IBM SPSS Modeler Premium Server. Using client/server architecture, SPSS Modeler Premium Server makes it possible for many analysts to work simultaneously without straining computing resources. The enterprise version supports in-database mining on leading information platforms and efficiently processes large amounts of data. SPSS Modeler Premium Server also offers additional deployment options to help you extend the benefits of data and text mining across geographic or functional lines and put results in the hands of decision makers quickly.

For additional IT efficiency, SPSS Modeler Premium Server can now be installed on IBM System z hardware running a Linux operating system. With greater computing power and the ability to handle multiple workloads simultaneously, such System z environments make it possible to analyze more data more rapidly – while meeting today’s needs for performance, scalability and reliability.

SPSS Modeler Premium Server also offers additional deployment options to help you extend the benefits of data and text mining across geographic or functional lines and put results in the hands of decision makers quickly. For example, to help manage your analytical assets and automate analytical processes, use SPSS Modeler Premium with IBM SPSS Collaboration and Deployment Services. You can also use insights derived from text data to achieve more accurate results with IBM SPSS Decision Management applications – for example, improve real-time and batch scoring, provide real-time recommendations to inbound callers or speed insurance claim processing.
### IBM SPSS Modeler 14.2 Premium – features

#### Data understanding
- Create a wide range of interactive graphs with automatic assistance
- Use visual link analysis to see the associations in your data
- Interact with data by selecting regions or items on a graph and viewing the selected information; or select key data for use in analysis
- Access SPSS Statistics graphs and reporting tools directly from the SPSS Modeler Premium interface

#### Data preparation
- Access operational data from Cognos Business Intelligence, IBM DB2, Oracle, Microsoft SQL Server, Informix®, Neoview, IBM Netezza, mySQL (Sun) and Teradata data sources, as well as mainframe data through z/DB2 and IBM Classic Federation Server support
- Import delimited and fixed-width text files, SPSS Statistics files, SAS, SPSS Data Collection data sources or XML
- Choose from the multiple data-cleaning options available in SPSS Modeler to remove or replace invalid data, automatically impute missing values and mitigate for outliers and extremes
- Apply automatic data preparation to interrogate and condition data for analysis in a single step
- Export data to delimited text files, Excel, SPSS Statistics, SAS, IBM Cognos Business Intelligence packages and operational databases
- Use field filtering, naming, derivation, binning, re-categorization, value replacement and field reordering
- Apply 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
- Choose from options for data restructuring, partitioning and transposition
- Select from extensive string functions: string creation, substitution, search and matching, whitespace removal and truncation
- Access data management and transformations performed in SPSS Statistics directly from SPSS Modeler
- Apply RFM scoring: aggregate customer transactions to provide Recency, Frequency, and Monetary value scores and combine these to produce a complete RFM analysis

#### Text-specific understanding and preparation features
- Extract text data from files, operational databases and RSS feeds (i.e., blogs, web feeds)
- Select native language extractor options for Dutch, English, French, German, Italian, Portuguese, Spanish or Japanese or translate virtually any language using third-party translation software
- Extract domain-specific concepts such as uniterms, expressions, abbreviations, acronyms and more
- Calculate synonyms using sophisticated linguistic algorithms and embedded or user-specified linguistic resources
- Name concepts by person, organization, term, product, location and other user-defined types
- Extract non-linguistic entities such as address, currency, time, phone number and Social Security number
- Use and customize pre-built templates and libraries for sentiment analysis, CRM, security and intelligence, market intelligence, life sciences and IT
- Leverage pre-packaged Text Analytics Packages (TAPs) for the most common business applications, or create your own
- Create clusters based on term co-occurrence using concept clustering algorithms, which provide an at-a-glance view of main topics and the way in which they are related
- Intelligently group text documents and records based on content, using text classification algorithms
- Enable advanced concept selection and deselection for use in predictive modeling
- Use text-based and visual reports to interrogate concept relationship, occurrence, frequency and type

#### Text link analysis
- Identify and extract sentiments (for example, likes and dislikes) from text in Dutch, English, French, German and Spanish
- Identify links and associations between, for example, people and events or diseases and genes
- Identify and extract content from URLs within blogs
- Include opinions, semantic relationships and linked events in deployable predictive models
- Reveal complex relationships through interactive graphs that show multiple semantic links between two concepts
| Modeling and evaluation | • 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 data attributes on predicted outcomes with variable importance graphs  
| | • Combine multiple models (ensemble modeling) or use one model to analyze a second model  
| | • Use automatic classification (binary and numeric) and clustering in place of selecting individual algorithms  
| | • Use the SPSS Modeler Component-Level Extension Framework (CLEF) to integrate custom algorithms  
| | • Through the integration of SPSS Statistics, use R to extend analysis options  
| Modeling algorithms included | • C&RT, C5.0, CHAID & QUEST – Decision tree algorithms including interactive tree building  
| | • Decision List – Interactive rule-building algorithm  
| | • K-Means, Kohonen, Two Step, Discriminant, Support Vector Machine (SVM) – Clustering and segmentation algorithms  
| | • Factor/PCA, Feature Selection – Data reduction algorithms  
| | • Regression, Linear, GenLin (GLM) – Linear equation modeling  
| | • Self-learning response model (SLRM) – Bayesian model with incremental learning  
| | • Time-series – Generate and automatically select time-series forecasting models  
| | • Neural Networks – Multi-layer perceptrons with back-propagation learning, and radial basis function networks  
| | • Support Vector Machine – Advanced algorithm with accurate performance for wide datasets  
| | • Bayesian Networks – Graphical probabilistic models  
| | • Cox regression – Calculate likely time to an event  
| | • Anomaly Detection – Detect unusual records through the use of a cluster-based algorithm  
| | • KNN – Nearest neighbor modeling and scoring algorithm  
| | • Apriori – Popular association discovery algorithm with advanced evaluation functions  
| | • CARMA – Association algorithm which supports multiple consequents  
| | • Sequence – Sequential association algorithm for order-sensitive analyses  
| | • IBM InfoSphere in-database mining algorithms supported*: Decision Tree, Association, Sequence, Regression, Logistic Regression, Clustering, Naïve Bayes, Time-Series and Radial Basis Function (RBF)  
| | • In-database algorithms for IBM Netezza*: IBM Netezza Decision Tree, IBM Netezza K-Means Clustering  
| | • Microsoft SQL Server in-database mining algorithms supported: Decision Tree, Association Rules, Linear Regression, Clustering, Sequence Clustering, Naïve Bayes, Time-Series and Neural Network  
| | • Oracle in-database mining algorithms supported: Decision Tree, General Linear Model (GLM), O-Cluster (Orthogonal Partitioning Clustering), KMeans, Apriori, Minimum Description Length (MDL), Support Vector Machine, Naïve Bayes, Adaptive Bayes, Non-Negative Matrix Factorization and Artificial Intelligence (AI)  
| Deployment | • Export models using SQL or PMML (the XML-based standard format for predictive models)  
| | • Leverage SPSS Collaboration and Deployment Services for innovative analytics management, process automation and deployment capabilities  
| Modeler server (optional) | • Use in-database mining to build models in the database using leading database technologies and leverage high-performance database implementations  
| | • Use SQL-pushback to push data transformations and select modeling algorithms directly into your operational databases  
| | • Leverage high-performance hardware, including IBM System z machines, to experience quicker time-to-solution and achieve greater ROI through parallel execution of streams and multiple models  
| | • Transmit sensitive data securely between SPSS Modeler Premium Client and SPSS Modeler Premium Server through secure sockets layer (SSL) encryption |
About IBM Business Analytics
IBM Business Analytics software delivers actionable insights decision-makers need to achieve better business performance. IBM offers a comprehensive, unified portfolio of business intelligence, predictive and advanced analytics, financial performance and strategy management, governance, risk and compliance and analytic applications.

With IBM software, companies can spot trends, patterns and anomalies, compare “what if” scenarios, predict potential threats and opportunities, identify and manage key business risks and plan, budget and forecast resources. With these deep analytic capabilities our customers around the world can better understand, anticipate and shape business outcomes.

For more information
For further information or to reach a representative please visit ibm.com/analytics.

Request a call
To request a call or to ask a question, go to ibm.com/business-analytics/contactus. An IBM representative will respond to your inquiry within two business days.

*Available with IBM SPSS Modeler Premium Server