

Improve Model Accuracy with Unstructured Data



Highlights

- Easily access, prepare and integrate structured data and text, Web and survey data
 - Support the entire data mining process with a broad set of tools based on CRISP-DM methodology
 - Quickly identify and extract sentiments from text in more than 30 languages and use this insight to build more accurate predictive models
 - Deploy textual insights so your entire organization benefits from a comprehensive, 360-degree view of the people you serve
-

IBM SPSS Modeler helps you achieve superior outcomes by basing your business decisions on patterns and associations found in your data. With Modeler, 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 – in comments, files or on the Web – modeling with structured data alone may provide an incomplete view into your business processes and outcomes. IBM SPSS Modeler Premium transforms the comprehensive Modeler workbench into a fully integrated data and text mining solution.

With Modeler Premium, you can combine all of your structured data - plus free-form text from documents, e-mails, call center notes, blogs, RSS feeds and other Web 2.0 sources – and apply the broad range of advanced data mining techniques available in Modeler Professional. By incorporating text sources into your modeling efforts, you can extract and discover relationships between concepts and sentiments, and increase your models' overall accuracy, or lift.

Boost productivity, quality and accuracy

With Modeler Premium, you can 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.



From this 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 – as well as textual data and data from Web 2.0 sources, such as RSS feeds, and IBM® SPSS® Data Collection products. No other data mining solution offers this versatility.

Modeler Premium's powerful automation tools, including automated data preparation and auto modeling, make 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.

Explore a broader range of data

Only with 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 Modeler toolbar, and reuse these valuable resources with other products and applications, including IBM® SPSS® Text Analytics for Surveys. The linguistic resources in 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

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, 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. Support for Language Weaver™ software, which automates human language translation, provides the ability to translate more than 30 languages into English, including Arabic, Chinese and Russian.

Optimize your current information technologies

Modeler Premium's open and scalable architecture 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

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, Modeler 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. Modeler 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.

Text insights deployed through the software's predictive models to operational databases provide value to areas throughout your organization.

To help manage your analytical assets and automate analytical processes, use Modeler Premium with IBM® SPSS® Collaboration and Deployment Services. You can also use insights derived from text data to achieve more accurate results with other IBM SPSS predictive applications - for example, improve real-time and batch scoring, provide real-time recommendations to inbound callers or speed insurance claim processing.

What's new in Modeler Premium 14

This release includes new features and enhancements that will enable you to create and interpret models easily using cutting-edge techniques, integrate seamlessly with other IBM SPSS software and technologies, and embed predictive modeling into your organization's business processes.

Performance improvements

- Improve the stability and accuracy of your models by leveraging large dataset optimization techniques, including boosting and bagging, for Neural Net, Linear and Decision Tree algorithms
- Run many models at once and interact with them using a new visualization tool that enables you to better understand the results of ensemble models and share them with others in your organization
- Enhance scalability and performance when using Modeler Premium Server by leveraging the new large database processing optimization capability for key algorithms. Build and refresh models on databases of unlimited size for enterprise-scale processes.

Updated algorithms

- The new Neural Net algorithm supports new analysis methods and includes multilayer perception and radial basis functions. It includes innovative, interactive visualization that makes it easy to understand and communicate results.
- Build better linear models using a new Linear Regression method that leverages large database processing, has built-in automatic data preparation options and produces rich visualizations that make it easy to interpret model results interactively

Text analysis enhancements

- 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
- Define and test rules on sample text before applying them to your data, using the enhanced text link rule editor

Data enhancements

- Extended support for enterprise data sources with the ability to read and write data to and from XML
- Leverage the strength of your operational databases better with additional in database mining options and push Modeler results back to operational database tables from the interface
- Simplify the re-use of modeling streams across users and ensure proper parameter settings with runtime parameter prompts
- Gain more control when exporting results to Excel by adding to an existing workbook or specifying where results should be placed in a spreadsheet

Improved platform support and deployment

- Improve deployment and scoring using a visual deployment definition that includes automatic model rebuilding, branching and model refresh capabilities
- Manage enterprise-wide login standards with new support for standard single sign-on technology (SSO)

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 IBM SPSS Statistics graphs and reporting tools directly from the Modeler interface

Data preparation

- Access operational data from a variety of operational data sources such as IBM DB2®, Oracle®, Microsoft SQL Server™, Informix®, Neoview, Netezza, MySQL (Sun) and Teradata.
- Import delimited and fixed-width text files, Statistics files, SAS, Data Collection data sources or XML
- Choose from Modeler's multiple data-cleaning options that 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, Statistics, SAS, 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 Statistics directly from 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 Language Weaver
- 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 Modeler's Component-Level Extension Framework (CLEF) to integrate custom algorithms
- Through the integration of 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 Machines – 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

Deployment

- Export models using SQL or PMML (the XML-based standard format for predictive models)
- Leverage 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
- 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 Modeler Client and Modeler Server through secure sockets layer (SSL) encryption

About SPSS, an IBM Company

SPSS, an IBM Company, is a leading global provider of predictive analytics software and solutions. The company's complete portfolio of products - data collection, statistics, modeling and deployment - captures people's attitudes and opinions, predicts outcomes of future customer interactions, and then acts on these insights by embedding analytics into business processes. IBM SPSS solutions address interconnected business objectives across an entire organization by focusing on the convergence of analytics, IT architecture and business process. Commercial, government and academic customers worldwide rely on IBM SPSS technology as a competitive advantage in attracting, retaining and growing customers, while reducing fraud and mitigating risk. SPSS was acquired by IBM in October 2009. For further information, or to reach a representative, visit www.spss.com.



© Copyright IBM Corporation 2010

SPSS Inc., an IBM Company Headquarters,
233 S. Wacker Drive, 11th floor
Chicago, Illinois 60606

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc., an IBM Company. © 2010 SPSS Inc., an IBM Company. All Rights Reserved.

IBM and the IBM logo are trademarks of International Business Machines Corporation in the United States, other countries or both. For a complete list of IBM trademarks, see www.ibm.com/legal/copytrade.shtml.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

Any reference in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.



Please Recycle