IBM SPSS Modeler

Deliver predictive intelligence to help make better decisions and achieve better outcomes

Highlights

- Access a variety of data sources such as flat files, databases, data warehouses, and Hadoop distributions to find hidden patterns in data

- Implement machine learning algorithms specially designed to efficiently process and model big data, integrating with Apache Spark for fast in-memory computing

- Deliver predictive, resource-aware and strategically aligned decisions to people and systems at the point of impact almost instantly

- Solve your business problems with a single platform that is designed to handle simple descriptive analysis all the way to the most complex optimization problems

- Analyze vast amounts of data in less time while fully using your existing IT investments with in-database performance and minimized data movement

- Take advantage of an open platform that can be deployed in most environments and integrated with other IBM solutions to bridge the gap between analytics and action

Overview

In a business environment, the main objective of analytics is to improve a business outcome. These outcomes can include:

- Increasing revenue by reducing customer attrition
- Increasing cross sell rates with a call center
- Decreasing costs by identifying fraudulent claims before payment
- Servicing a component in a production line to minimize downtime

When you apply analytics to improve a decision, the result is likely to be a better outcome.

Data mining is the process of using analytical techniques to uncover patterns in data. Descriptive analysis, predictive modeling, text analytics, geospatial analytics, entity analytics, decision management and optimization are used to identify patterns and deploy predictive models into operational systems. Systems and people can use these patterns and models to derive insights that enable them to consistently make the right decision at the point of impact. Outcomes are maximized based on the predictive intelligence hidden in data of growing size and complexity.

IBM® SPSS® Modeler is a powerful predictive analytics platform that is designed to bring predictive intelligence to decisions made by individuals, groups, systems and your enterprise. SPSS Modeler scales from desktop deployments to integration with operational systems to provide you with a range of advanced algorithms and techniques. Applying these techniques to decisions can result in rapid ROI and can enable organizations to proactively and repeatedly reduce costs while increasing productivity.

SPSS Modeler is available in four editions to meet virtually any analytical needs:

IBM SPSS Modeler Professional

Uncover hidden patterns in structured data. SPSS Modeler Professional provides advanced algorithms, data manipulation and automated modeling and preparation techniques to build predictive models that can help you deliver better business outcomes.

IBM SPSS Modeler Premium

Unlock insights from almost any type of data using a range of advanced algorithms and capabilities, such as text analytics, entity analytics and social network analysis, along with automated modeling and preparation techniques.

IBM SPSS Modeler Gold

Build and deploy predictive models directly into your business processes with decision management capabilities so people and systems can make the right decisions every time. With SPSS Modeler Gold on Cloud, these capabilities are available as a web-based monthly subscription service.

IBM SPSS Predictive Analytics Enterprise

Optimize your decision-making at the point of impact with a single, multifaceted predictive analytics solution that includes SPSS Modeler.

With each edition of SPSS Modeler, you can:

- Access all types of data.
- Broaden your analytics reach.
- Accommodate your needs with flexible deployment.

Accessing all types of data

Data is being generated at an exponential rate from a multitude of sources, thereby fueling new information and untapped opportunities for those organizations able to harness it and realize its value. This data is stored in various systems and formats so bringing it together can be a challenge. The volume of data is so big that you cannot analyze it manually nor can you look over tables in reports to find why something might or might not happen. The analysis process presents yet another challenge because of a scarcity of skilled analysts that can work with the data to extract its value.

With SPSS Modeler, you and your organization can use the data you have available—both spatial and non-spatial—and extract value from it by discovering untapped opportunities and new information. With new insights from your data, you can predict what is likely to happen, become proactive and optimize outcomes, rather than reacting simply as your current situation dictates.

SPSS Modeler enables you to use a range of analytical techniques to access data sources such as flat files, databases, data warehouses, and Hadoop distributions. These statistical techniques use historical data to make predictions about current conditions or future events. Also included are capabilities for data access, data preparation, data modeling and interactive visualizations. With automated procedures for preparation and modeling, it is suitable for a wide range of analytics abilities.

The intuitive graphical interface of SPSS Modeler enables users to visualize each step of the data mining process as part of a "stream." By interacting with these streams, analysts and business users can collaborate, which adds business knowledge and domain expertise to the data mining process. Users can focus on discovering insights rather than on technical tasks such as writing code. They can also pursue "train-of-thought" analysis and explore data more deeply, both of which can uncover additional relationships that make sense to your organization.

Data preparation and manipulation

Preparing data for analysis is an important but time-consuming step in analysis. SPSS Modeler automates data preparation to ease the process and to help you make sure your data is in the best format for analysis. The tasks automated include analyzing data and identifying fixes, screening out fields, deriving new attributes when appropriate, and improving performance through intelligent screening techniques.

SPSS Modeler offers a variety of ways to manipulate and prepare data for analysis at the record or field (or variable) level. Among the methods used to help make sure your data is in the best format for the specific type of analysis that is being undertaken are:

- *Record operations.* Select, Sample and Distinct nodes enable you to choose specific rows of data. You can merge and append nodes to join data by adding columns or rows to a dataset. Aggregate and Recency, Frequency, Monetary (RFM) Aggregate nodes summarize records to a single row. A Balance node adjusts the proportions of records in imbalanced data and a Sort node reorders based on value. The Space Time Box node creates geospatial and time-based data for records.
- *Field operations.* A Type node specifies metadata and properties of a dataset, and the Filter node discards fields. The Derive node creates new fields and a Filler node can replace existing field values. Data can be restructured with the Set to Flag, Restructure or Transpose nodes and regrouped with the Reclassify or Binning nodes. To assist with modeling, the Partition node can split the data and the History node and Time Intervals nodes can create additional fields. The Field Reorder node defines the display ordering to make certain fields easier to view.

Broadening your analytics reach with a range of techniques

Analytical techniques are continuing to evolve, providing analysts with a plethora of options for tackling the problems in front of them. Additionally, as technology develops and new types of data become available (such as location-based data from mobile phones or cell towers), different questions and challenges arise about the best ways to exploit this data. Innovative techniques are therefore necessary.

With SPSS Modeler, your analysts can solve their business problems with a single platform that is designed to handle simple descriptive analysis, the most complex optimization problems—and everything in between. SPSS Modeler features capabilities that go beyond the standard analytic requirements of today's analysts. A range of models, automated modeling and data preparation, text analytics, entity analytics, social network analysis and the ability to build models on parallel processes can all help you address the most sophisticated problems.

Automated data modeling

With the automated modeling features of SPSS Modeler, non-analysts can produce accurate models quickly without specialized skills. In addition, advanced predictive modeling capabilities enable professional analysts to create the most sophisticated of streams.

Automated modeling enables you to compare multiple modeling approaches. By setting specific options for each model type (or using the defaults), you can explore a multitude of model combinations and options. The generated models are then ranked based on the measure specified, saving the best for use in scoring or further analysis.

A range of models

SPSS Modeler offers an array of modeling techniques, including all of the following algorithms:

- *Classification algorithms*. Make predictions or forecasts based on historical data with techniques. Examples include decision trees, random trees, neural networks, logistic regression, support vector machines, Cox regression, generalized linear mixed models (GLMM) and more. Use automatic classification modeling for both binary and numeric outcomes to streamline model creation or Self-Learning Response Modeling (SLRM) to build a model that you can continually update or re-estimate without having to rebuild the model.
- *Segmentation algorithms*. Group people or detect unusual patterns with techniques such as automatic clustering, anomaly detection and clustering neural networks. 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 with Apriori, CARMA and sequential association.
- *Time series and forecasting*. Generate forecasts for one or more series over time with statistical modeling techniques. Using temporal causal modeling, you can discover causal relationships among a large number of series.
- *Extendibility with R and Python programming languages*. Apply transformations, use scripts to analyze, summarize or produce text and graphical output with R. With the Custom Dialog Builder, you can share and reuse R and Python and Python code with those who choose not to use programming for analysis.
- *Monte Carlo simulation*. Account for uncertainty in inputs to predictive models. Model uncertain inputs based on historical data or with probability distributions to generate simulated values, and then use them in the predictive model to generate an outcome. The result is a distribution of outcomes that can provide answers to questions that are based on realistically generated data.

- *Entity analytics*. Identify relationships and improve the coherence and consistency of current data by resolving identity conflicts in the records themselves. Identifying these relationships can be vitally important in a number of fields, including customer relationship management, fraud detection, anti-money laundering and security.
- *Social network analysis*. Discover the relationships between social entities and the implications of these relationships on an individual's behavior. These capabilities can help transform information about relationships into key performance indicators that show the social behavior of individuals and groups.

Geospatial analytics

With SPSS Modeler, you can explore the relationship of data elements that can be tied to a location and perform geographic spatial analysis of your data to reveal insights that would not be visible in charts or tables. With spatial mining, you can easily mine geospatial data using ESRI shape files. By analyzing both non-spatial and spatial data, overall model accuracy is improved and you are able to gain deeper insights into people and events.

Add a new dimension to your analysis by discovering association rules among spatial and non-spatial attributes. Using spatial temporal prediction, you can fit linear models for measurements taken over locations in 2D space, enabling you to easily predict "hot" areas and how those areas will change over time. You can apply this technology to mine geospatial data in fields such as crime pattern analysis, epidemic surveillance, building management and branch performance analysis.

Text analytics

The interactive, visual environment of SPSS Modeler uses advanced linguistic technologies and Natural Language Processing (NLP) to rapidly process unstructured text data. From this text, it extracts and organizes the key concepts. The customizable industry-specific text analysis packages enable you to analyze relevant terms and phrases in addition to acronyms, emoticons and slang in the right context. Interactive graphs help you explore and display text data and patterns for instant analysis. You can create hierarchical categorization structures and include them as inputs to a predictive model to yield better and more focused decisions and results. Predefined categories, such as hierarchies, annotations and keyword descriptors can be imported to categorize initial unstructured data so you can organize concepts more logically and in greater detail.

Accommodating your needs with flexible deployment

The deployment of analytics in your organization will depend on many environmental factors. Such factors include the business problems that must be addressed, your choice of operating systems and platforms, and the other technologies and data sources in your infrastructure. Technology, and particularly software, should be flexible enough to accommodate various permutations and still provide the expected performance and results.

The SPSS Modeler architecture is an open one that supports a range of platforms and languages.

You can deploy SPSS Modeler locally, from the cloud, or even using a hybrid cloud approach, then confidently use it with your existing systems to help optimize performance and address your business problems. This flexible deployment bridges the gap between analytics and action by providing results to people and processes on a schedule or on demand. SPSS Modeler streams can be deployed for the purposes of model refresh or automated job scheduling. Developers can embed SPSS Modeler streams into Java applications using the embedded predictive modeling (EPM) Java API. And you can deploy them with decision management or other predictive applications.

Decision management

Decision management extends the predictive capabilities of SPSS Modeler to everyday business processes to empower front-line employees and systems. It integrates predictive models, simple rules and scoring into your systems to automate, manage and optimize high-volume decisions. It then recommends actions where and when people need them, such as cross-selling on the phone with a customer, deciding the best routing for a claim, using a utility to allocate bandwidth or presenting offers in a self-service kiosk. Thousands of decisions can be made at the operational level in complete alignment with your organization's goals and strategies.

With the decision management capability of SPSS Modeler:

- Predictive models can foresee the most likely outcomes and identify the factors driving the outcome, such as the propensity of a customer to respond to a given offer or the risk that a given claim is fraudulent.
- Business rules automate parameters that are determined by elements such as business policies or legal and regulatory compliance. Basic rule support is provided directly in SPSS Modeler. For more robust rules that scale to meet enterprise-wide requirements, integration with IBM Operational Decision Management is also supported.
- Integrated scoring can provide almost instant recommendations to the right people and systems so that resource-aware and strategically-aligned decisions can be made, no matter the line of business.

Optimization

Even the most committed organization cannot afford to move every customer to the front of the line, nor can an insurer investigate every claim. Businesses operate under real-world constraints, subject to limits on available staff, equipment and investment. Optimization enables organizations to make the most of scarce resources by identifying the solution that best meets a specific goal. Examples include maximizing the revenue from a marketing campaign or minimizing the risk of fraud or churn. Optimization can be used on demand, such as on an individual to determine the best offer for that person, or in batch to allocate offers to all eligible customers.

Execution and scheduling

SPSS Modeler includes capabilities that are designed to use automation to bring greater consistency to your results. Greater consistency strengthens people's confidence in analytics because management can efficiently govern the business environments where analytical processes take place. This governance helps ensure that all internal and external procedural requirements are met.

With SPSS Modeler, your analysts can construct flexible, repeatable analytical processes that can be operationalized, that is, initiated at the right time and integrated with other enterprise processes. Predefined model management processes help models remain relevant and accurate.

In-database

SPSS Modeler provides a number of capabilities to minimize data movement and push analytics to the database, such as:

- *SQL Pushback*. With SPSS Modeler Server, moving data from large databases, even in IBM® System z® and IBM PureSystems® environments, is not required because the analytics and mining can take place in the database. SQL Pushback enables in-database data transformation and preparation without the need to write any SQL or do any programming. The result is a significant improvement in analytical performance.
- *In-database scoring*. Database-specific scoring adapters, which are available for IBM SPSS Modeler with Scoring Adapter for zEnterprise®, IBM DB2®, IBM PureData® System for Analytics (powered by Netezza®) and Teradata solutions, extend the number of SPSS Modeler algorithms that can be scored in database, further reducing the need to extract the data before scoring.
- *Database functions*. SPSS Modeler can use almost any user-defined functions (UDFs), database aggregate and windowed aggregate functions provided by a database. These are exposed by the SPSS Modeler workbench to extend the available native functionality and ensure SQL Pushback.

- *In-database mining*. SPSS Modeler Server supports integration with the data mining capabilities, modeling tools and database-native algorithms that are available with IBM DB2 Analytics Accelerator (IDAA) on Z " Hytap" (Hybrid Transactional and Analytic Processing), PureDataSystem for Analytics, Oracle Data Miner, Microsoft Analysis Services and others. You can build, score, and store models inside the database—all from the SPSS Modeler workbench.

Integration with IBM technology

SPSS Modeler includes capabilities for exporting data to IBM Cognos® Business Intelligence and Cognos TM1® software. The results of analysis can be distributed for reporting, monitoring and planning to key decision-makers who only need the results. When further analysis is needed, SPSS Modeler can also access them as a data source, which means the process can continue again, thereby feeding the results back to wherever the questions began.

IBM SPSS Statistics provides the ability to carry out further statistical analysis and data management to complement SPSS Modeler and its data mining abilities with a dedicated section on the nodes.

SPSS Modeler provides support for PureData System for Analytics to access specific models from the SPSS Modeler Interface and leverage the hardware's speed and performance.

For analytics on big data, the addition of SPSS Analytic Server to SPSS Modeler enables analytics to be processed in a Hadoop distribution.

Integration of SPSS Modeler with Apache Spark through SPSS Analytic Server provides a broad range of big data algorithms and in-memory processing capabilities for creation of fast and efficient machine learning models.

SPSS Modeler streams can also be deployed with InfoSphere® Streams for high-velocity streaming data applications requiring predictive scores.

Conclusion

SPSS Modeler is a predictive analytics platform that scales from desktop deployments to integration in operational systems to bring predictive intelligence to decisions made by individuals, groups, systems and the enterprise. Your organization can use SPSS Modeler to conduct analysis regardless of where the data is located, the size of the data, or whether it is structured or unstructured. The scalable client-server architecture enables users to access everything from flat files to big data environments. Analysis is pushed back to the source for execution, minimizing data movement and increasing performance.

With SPSS Modeler, all kinds of users can solve a variety of business problems. It offers analytics techniques that range from descriptive analytics to advanced algorithms, including automated modeling, text analytics, entity analytics, social network analysis, decision management and optimization. An intuitive interface is designed for a wide range of users from the non-technical business user to the analytical professional. The short learning curve for SPSS Modeler makes it appealing to the novice and advanced user, so they can quickly uncover insights and realize real business results.

Learn more

Explore more predictive analytics software and resources at: ibm.biz/predictive.

Take advantage of our active and growing open-source community, where you can find resources to help you expand the use of IBM predictive analytics software. Resources include blogs, videos, tutorials, and an extensive library of more than 6,000 predictive extensions to help you take advantage of popular programming languages such as R, Python and Java. Join the community at: developer.ibm/predictiveanalytics

About IBM Analytics

IBM Analytics software delivers data-driven insights that help organizations work smarter and outperform their peers. This comprehensive portfolio includes solutions for business intelligence, predictive analytics and decision management, performance management, and risk management.

IBM Analytics solutions enable companies to identify and visualize trends and patterns in areas, such as customer analytics, that can have a profound effect on business performance. They can compare scenarios, anticipate potential threats and opportunities, better plan, budget and forecast resources, balance risks against expected returns and work to meet regulatory requirements. By making analytics widely available, organizations can align tactical and strategic decision-making to achieve business goals. For further information please visit **ibm.com**/analytics

Request a call

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

YTD03341-USEN-04