



# Data Mining From A to Z:

How to Discover Insights and Drive Better Opportunities

# Contents

Introduction .....	1
The SAS® Analytical Life Cycle: Combining Data, Discovery and Deployment.....	2
What Can Data Mining Help You Discover? .....	3
A Closer Look at the Role of Data Mining in the Discovery Process .....	5
Step 1: Turn a Business Question Into an Analytical Hypothesis .....	5
Step 2: Prepare the Data for Data Mining .....	5
Step 3: Explore the Data .....	5
Step 4: Model the Data .....	6
SAS® Data Mining Solutions .....	6
Using SAS® Enterprise Miner™ for Data Mining and Machine Learning .....	6
Using SAS® Factory Miner for an Automated Approach to Data Mining .....	8
Scaling Your Discovery Process to Handle Big Data and Complex Problems .....	8
Integration Eases Model Deployment, Monitoring and Management.....	10
Conclusion.....	11
Learn More .....	11

## Introduction

So much data and multitudes of decisions. Organizations everywhere struggle with this dilemma. The data is growing, but what about your ability to make decisions based on those huge volumes of data? Is that growing too? For many, unfortunately, the answer is no.

Data pours in at unprecedented speeds and volumes from everywhere. But making fact-based decisions is not dependent on the amount of data you have. Actually, having so much data can be a paralyzing factor. Where do you even begin? Your success will depend on how quickly you can discover insights from all that data and use those insights to drive better actions across your entire organization.

That's where predictive analytics, data mining, machine learning and decision management come into play. Predictive analytics helps assess what will happen in the future. Data mining looks for hidden patterns in data that can be used to predict future behavior. Businesses, scientists and governments have used this approach for years to transform data into proactive insights. Decision management turns those insights into actions that are used in your operational processes. So while the same approaches can still be applied today - they need to happen faster and at a larger scale, using the most modern techniques available.

Forward-thinking organizations use data mining and predictive analytics to detect fraud and cybersecurity issues, manage risk, anticipate resource demands, increase response rates for marketing campaigns, generate next-best offers, curb customer attrition and identify adverse drug effects during clinical trials, among many other things.

Because they can produce predictive insights from large and diverse data, the technologies of data mining, machine learning and advanced analytical modeling are essential for identifying the factors that can improve organizational performance and, when automated in everyday decisions, create competitive advantage. And with more of everything these days (data, computing power, business questions, risks and consumers), the ability to scale your analytical power is essential for staying ahead of your competition.

Deploying analytical insights quickly ensures that the timeliness of your analytical models is not lost due to slow processes like rewriting code for each environment, revalidating the rewritten models and other manual processes. If you can rapidly deploy your analytical models, the context and relevance of the models is not lost and you retain competitive advantage.

So how do you create an environment that can help your organization deal with all of the data being collected, all of the models being created and all of the decisions that need to be made, all at an increasing scale? The answer is an iterative analytical life cycle that brings together:

- **Data** - the foundation for decisions.
- **Discovery** - the process of identifying new insights in data.
- **Deployment** - the process of using newly found insights to drive improved actions.

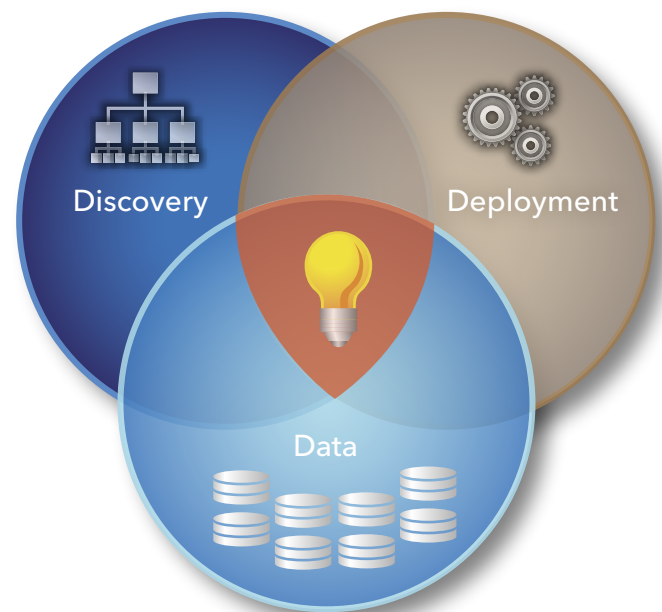


Figure 1: An integrated combination of data, discovery and deployment is needed to derive and put into action the fast insights needed for scalable decision making.

## The SAS® Analytical Life Cycle: Combining Data, Discovery and Deployment

Even though the majority of this paper is focused on using data mining for insights discovery, let's take a quick look at the entire iterative analytical life cycle, because that's what makes predictive discovery achievable and the actions from it more valuable.

- **Ask a business question.** It all starts here. The discovery process is driven by asking business questions that produce innovation. This step is focused on exploring what you need to know, and how you can apply predictive analytics to your data to solve a problem or improve a process.
- **Prepare data.** Collecting data certainly isn't a problem these days – it's streaming in from everywhere. Technologies like Hadoop and faster, cheaper computers have made it possible to store and use more data, and more types of data, than ever before. But there is still the issue of joining data in different forms from different sources and the need to transform raw data into data that can be used as input for data mining. Data scientists still spend much of their time dealing with these tasks.
- **Explore the data.** Interactive, self-service visualization tools need to serve a wide range of user personas in an organization (from the business analyst with no analytical knowledge to a data scientist) to allow searches for relationships, trends and patterns to gain deeper understanding of the information captured by variables in the data. In this step, the hypothesis formed in the initial phase of the project will be refined and ideas on how to address the business problem from an analytical perspective are developed and tested. While examining your data, you may find the need to create, select or transform some data to create more precisely focused models. Fast, interactive tools help make this an iterative process, which is crucial for identifying the best questions and answers.
- **Model the data.** In this stage, the data scientist applies numerous analytical modeling algorithms to the data to find a robust representation of the relationships in the data that help answer the business question. Analytical tools search for a combination of data and modeling techniques that reliably predict a desired outcome. Experimentation is key to finding the most reliable answer, and automated model building can help minimize the time to results and boost the productivity of analytical teams. In the past, with manual model-building tools, data miners and data scientists were able to create several models in a week or month. Today, they can create hundreds or even thousands. But how can they quickly and reliably find the one model (out of many) that performs best? With automated tournaments of machine-learning algorithms and a clearly defined champion model, this has become an easy process. Analysts and data scientists can now spend their time focusing on more strategic questions and investigations.
- **Implement the models.** Here we move from the discovery phase to deployment – taking the insights learned and putting them into action using repeatable, automated processes. In many organizations this is the point where the process often slows down dramatically because there is no defined handshake between the two worlds of discovery and deployment, let alone automation. Bringing these two worlds together to create an integrated transition helps decrease time to value for predictive analytics. The faster your business can use the answers generated by predictive analytics for better decision making, the more value will be generated. And, a transparent process is important for everyone – especially auditors.
- **Act on the new information.** There are two types of decisions that can be made based on analytical results. Strategic decisions are made by humans who look at results and take action. Operational decisions are automated – like credit scores or recommended best offers – and don't require human intervention. More and more organizations are looking to automate operational decisions and provide real-time answers and results to reduce decision latencies. Basing operational decisions on answers from analytical models also makes decisions objective, repeatable and measurable. The integration with enterprise decision management tools enables organizations to build comprehensive and complete operational decision flows that combine data-driven analytics and business rules for optimal automated decisions.
- **Evaluate your results.** The next – and perhaps most important – step is to evaluate the outcome of the actions produced by the analytical model. Did your predictive models produce tangible results, such as increased revenue or decreased costs? With continuous monitoring and measurement of the models' performance, you can evaluate the success of these assets and make sure they continue to produce the desired results.



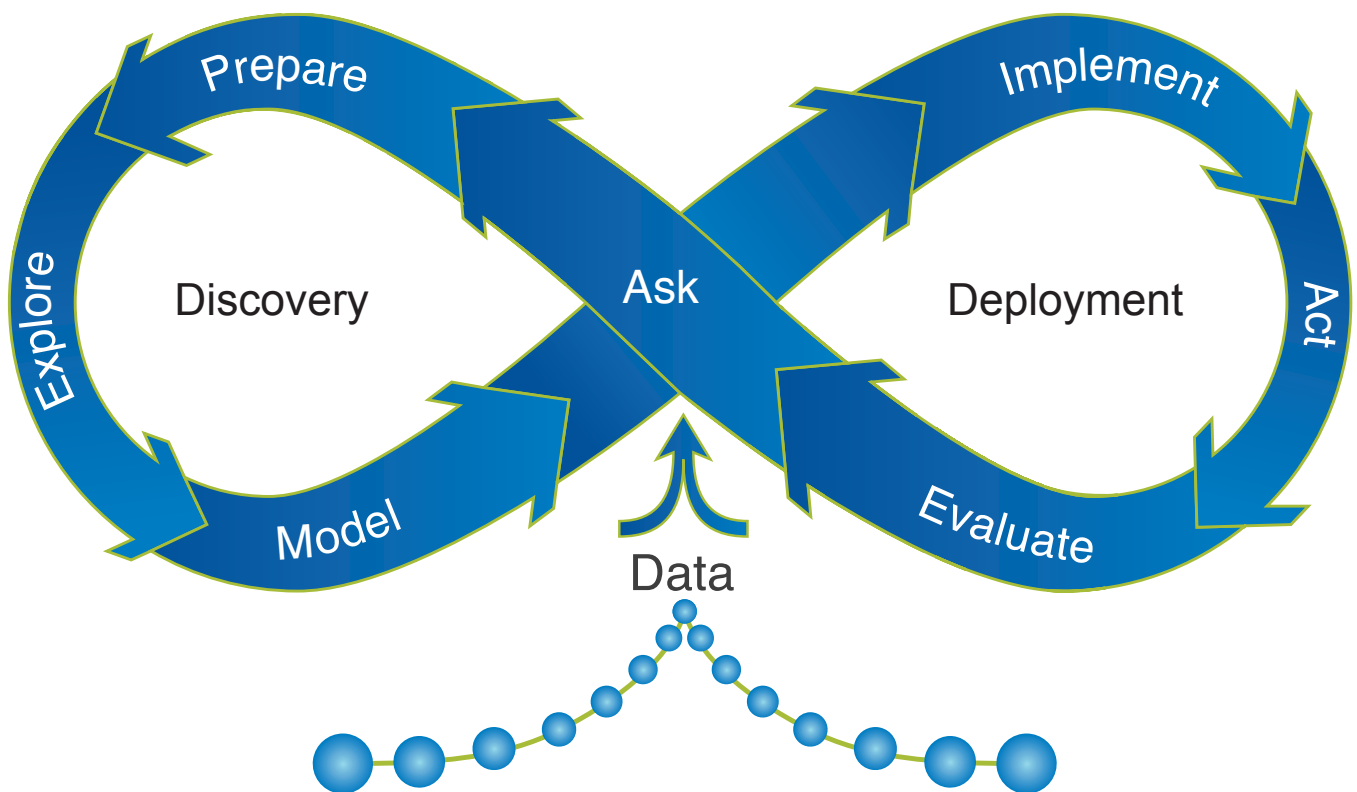


Figure 2: The analytical life cycle is an iterative process of making discoveries from your data and applying new insights to continually improve predictive models and their results.

- **Ask again.** Because your data is always growing and ever changing, relationships in data that your models use for predictions also change over time. Constant evaluation of your analytical results will identify the degradation of model accuracy. Even the most accurate models will have to be refreshed over time, and organizations will need to go through the discovery and deployment steps again. It's a constant and evolving process.

SAS provides an integrated, complete analytics platform that handles every step in the iterative analytical life cycle. This remainder of this paper will focus on the data discovery portion of the life cycle – and the data mining tools you'll need to quickly build the most accurate predictive models possible.

## What Can Data Mining Help You Discover?

Data mining provides a core set of technologies that help organizations anticipate future outcomes, discover new opportunities and improve business performance. It can be applied to a variety of customer issues in any industry – from customer segmentation and targeting, to fraud detection and credit risk scoring, to identifying adverse drug effects during clinical trials.

A common use of data mining and machine-learning techniques is to automatically segment customers by behavior, demographics or attitudes – to better understand needs of specific groups and serve them in a more targeted way. This analytical segmentation, or unsupervised modeling, helps to identify groups of customers that are similar and might react to

certain offers or activities in a similar way. Using these segments, you can create models for each group to predict the next-best offer or activity to which they're most likely to respond. To ensure that you only engage desired customers, you can further complement the customer acquisition model with a risk-scoring model to find out who is a good credit risk and actually worth the investment to acquire or retain.

Another important use for data mining and machine learning is to help detect fraud, which is important as fraudsters become more sophisticated in their tactics. Models can be built to cross-reference data from a variety of sources, correlating nonobvious variables with known traits to identify new patterns of fraudulent activities.

Because of its potential to produce accurate predictive insights from huge volumes of diverse data, data mining has proven to be an invaluable component of many analytical initiatives. Data mining and machine learning can help you:

- Automatically discover patterns, trends and relationships represented in data.
- Develop models to better understand and describe characteristics and activities based on these patterns.
- Use those insights to help evaluate future options and make fact-based decisions.
- Create score code that expresses the calculations to be made for timely, appropriate actions.

### Common Applications for Data Mining Across Industries

Business Question	Application	What Is Predicted?
How to better target product/service offers?	Profiling and segmentation.	Customer behaviors and needs by segment.
Which product/service to recommend?	Cross-sell and up-sell.	Probable customer purchases.
How to grow and maintain valuable customers?	Acquisition and retention.	Customer preferences and purchase patterns.
How to direct the right offer to the right person at the right time?	Campaign management.	The success of customer communications.
Which customers to invest in and how to best appeal to them?	Profitability and lifetime value.	Drivers of future value (margin and retention).

### Industry-Specific Data Mining Applications

Business Question	Application	What Is Predicted?
How to assess and control risk within existing (or new) consumer portfolios?	Credit scoring (banking).	Creditworthiness of new and existing sets of customers.
How to increase sales with cross-sell/up-sell, loyalty programs and promotions?	Recommendation systems (online retail).	Products that are likely to be purchased next.
How to minimize operational disruptions and maintenance costs?	Asset maintenance (utilities, manufacturing, oil and gas).	The real drivers of asset or equipment failure.
How to reduce health care costs and satisfy patients?	Health and condition management (health insurance).	Patients at risk of chronic, treatable/preventable illness.
How to decrease fraud losses and lower false positives?	Fraud management and cybersecurity (government, insurance, banks).	Unknown fraud cases and future risks.
How to bring drugs to the marketplace quickly and effectively?	Drug discovery (life sciences).	Compounds that have desirable effects.

## A Closer Look at the Role of Data Mining in the Discovery Process

Data mining and machine learning lie at the heart of the discovery process. But there's more to discovery than just building an analytical model. You'll get better results if you take an iterative, holistic approach.

### Step 1: Turn a Business Question Into an Analytical Hypothesis

The first step in the discovery process is to ask a business question (see tables on page 4). Usually an organization has a general idea of what it wants to achieve – something like, “We want to reduce the churn of our valuable customers.” To address these issues with analytics, business questions must be specified in detail or transformed into an analytical hypothesis. For example, every predictive model requires a well-defined outcome, a label or target. If you want to predict customer churn, you need to define churn as an outcome for the model. However, churn is likely defined differently in different organizations. Does it refer to someone actively canceling a contract or someone who is dormant in his activities? How long can a customer remain dormant before being classified as a churner? What is valuable? Do we include only historical value or potential future value (lifetime value) of a customer? Your first step in the discovery process is to identify an issue and translate the issue into a question that can be addressed with analytics.

### Step 2: Prepare the Data for Data Mining

To begin, you must determine what data is needed to answer the question. Based on the specifics of the business question, an analyst evaluates the data that is available and decides if the data has the potential to answer the question at hand. If not, external data may be needed or new data might need to be collected. Often, the data is in different systems and needs to be accessed and turned into a data set that can be used for data mining and machine learning. Predictive or supervised models require a single record per entity to model. (An analytics base table for forecasting or market analysis will look different from a table for predictive or supervised modeling). If you want to model the likelihood of customer churn, you need to create a single table where each record contains all the data attributes for a single customer. This often requires a significant amount of data aggregation and transformation. Once a single analytics base table for the analysis has been aggregated, the other aspects of the life cycle come into play. Because it is necessary to experiment with data, the preparation stage is also very iterative with the analyst trying different types of data to get the most accurate predictive results.

In some cases to expedite modeling processes, you may want to sample the data – that is, create a smaller subset of the data that represents the target data set. Data mining can only uncover patterns already present in the data, so the sample should be representative and large enough to contain the significant information. The analytics base table is also generally divided into at least two sets: the training set and the test set. The training set is used to train the data mining and machine-learning algorithm(s), while the test set is used to verify the accuracy of any patterns found.

### Step 3: Explore the Data

Next, you'll want to explore the data and search for anticipated relationships, unanticipated trends and anomalies to gain an understanding of the information you're working with and further refine ideas and questions. Data exploration can also help pinpoint data quality problems such as data errors, missing values or data distributions that need to be transformed for the modeling stage. In addition, you can use several other types of techniques to detect patterns in the data that can help you build more accurate predictive models or help you create additional input data for your predictive model.

- **Clustering (or unsupervised modeling)** identifies groups or structures in the data that are similar, beyond the structures otherwise visible in the data.
- **Association-rule learning** searches for relationships among variables, such as products frequently bought together (known as market basket analysis), which can lead to further recommendations for purchase.
- **Text analytics** can help you to create new structured information from electronic text data. This new data can help to improve the accuracy of your models. For example, integrating customer comments on your products and services from call center notes or reviews on social media forums often produces more accurate churn prediction models.
- **Interactive data visualization** presents results graphically and lets users interact with these graphs to more easily identify important patterns or anomalies with the data that might have an impact in the model-building stage.

Often, you'll need to modify your data before modeling so you should plan on a step for creating, selecting and transforming variables to focus your model-selection process. Based on your discoveries in the exploration phase, you may need to manipulate your data to introduce new variables, fill in missing values or look for outliers so you can reduce the number of variables to only the most significant ones.

## Step 4: Model the Data

After carefully exploring and preparing your input data, you are ready to create predictive or supervised models to search for a combination of the data that reliably predicts a desired outcome. Depending on the data and issue at hand, you can choose from a variety of modern machine-learning and statistical techniques to solve your problem – including classification, regression, neural networks, random forests, support vector machine, incremental response or time series data mining – as well as industry-specific techniques such as credit scoring in banking or rate making for insurance.

The selection of the most appropriate techniques depends on several factors. Is it more important to have a model that predicts your desired outcome with the highest accuracy or is it also (or even more) important to have transparency into the data relationships that drive the predictions? Automated machine-learning techniques are often too complex to allow the exploration of business drivers from the model results, while other statistical techniques such as regression or decision trees are more transparent and are preferred in regulated industries.

To get the most value from your predictive models, you'll want to constantly evaluate the usefulness and reliability of the findings from your data mining processes. Not all patterns found by the data mining algorithms will be valid. The algorithms might find patterns in the training data set that are not present in the general data set. (This is called overfitting.) To address this concern, patterns are validated against a test set of data. The patterns learned on the training data will be applied to the test set, and the resulting output is compared to the desired (or known) output.

For example, a data mining algorithm that had been trained to distinguish fraudulent credit card transactions from legitimate ones would then be applied to the test set of transactions on which it had not been trained. The accuracy of the patterns can then be measured from how many credit card transactions are correctly classified. If the learned patterns do not meet desired standards, modifications are made to the preprocessing and data mining techniques until the result is satisfactory and the learned patterns can be successfully applied to operational systems.

Data scientists and data miners need to experiment with a multitude of predictive modeling and machine-learning algorithms in order to find the one that works best for their specific problem. Automated modeling tournaments where users can experiment to identify the winning modeling strategy quickly

are big timesavers here. When you are satisfied with the results of your modeling endeavors, you then begin the deployment process. But because it's a completely iterative process, there are constant examinations and adjustments. As discussed before, there are several steps involved in the deployment process (see the SAS Analytical Life Cycle section on page 2). For more information on the deployment process, read [From Data to Decision: How SAS® Decision Manager Automates Operational Decisions](#). To learn more about data mining and discovery, keep reading!

## SAS® Data Mining Solutions

Data mining and machine learning enable you to discover insights that drive better decision making. With SAS data mining solutions, you can streamline the discovery process to develop models quickly so you can understand key relationships and find the patterns that matter the most.

### Using SAS® Enterprise Miner™ for Data Mining and Machine Learning

SAS Enterprise Miner is a comprehensive, graphical workbench for data mining. This widely acclaimed and extensive platform provides capabilities to prepare data for predictive analytics, identify the most significant variables, develop models using the most modern data mining and machine-learning algorithms, easily validate the accuracy and fitness of the model(s), and generate assets that allow a simple deployment of analytical models into your operational applications for automated decision making.

Powerful data preparation tools address data quality problems, such as missing values and outliers, and help you develop segmentation rules. Interactive data exploration enables users to create dynamic, linked plots to identify relationships within the data. SAS Enterprise Miner provides dozens of advanced statistical and machine-learning algorithms for descriptive and predictive modeling, including clustering, link and market basket analysis, principal component analysis, decision trees, bagging and boosting, Bayesian networks, neural networks, random forests, linear regression, logistic regression, support vector machine, time series data mining and many more.

At the end of the model development pipeline, complete, optimized scoring code is delivered for easy deployment of the unsupervised or supervised models in SAS, C, Java and PMML for scoring data in SAS as well as in other environments. Score code can also be delivered automatically as an in-database



function for scoring inside Hadoop as well as industry-leading databases such as Teradata, IBM, Oracle, Pivotal, Aster Data, SAP HANA, etc., for very seamless integration with business applications and fast operational results.

In addition to generating score code in different languages and formats, SAS Enterprise Miner also generates many assets that enable easy deployment, management and monitoring of predictive models as part of operational business processes. All of these assets are supported by metadata to provide meaningful documentation around the entire process.

The SAS Enterprise Miner data mining process is driven by a process flow diagram that you can modify, save and share. The drag-and-drop GUI enables business analysts with little statistical expertise to navigate through the data mining process, while the quantitative expert can go behind the scenes to fine-tune the analytical models.

### With SAS Enterprise Miner, you can:

- Create training and test sample data sets with high predictive value.
- Interactively explore relationships and anomalies in the data.
- Create, transform and select the most appropriate variables for analysis.
- Apply a range of modeling techniques to identify patterns in the data.
- Validate the usefulness and reliability of findings from the data mining process.
- Create all required assets for easy model deployment, monitoring and management.

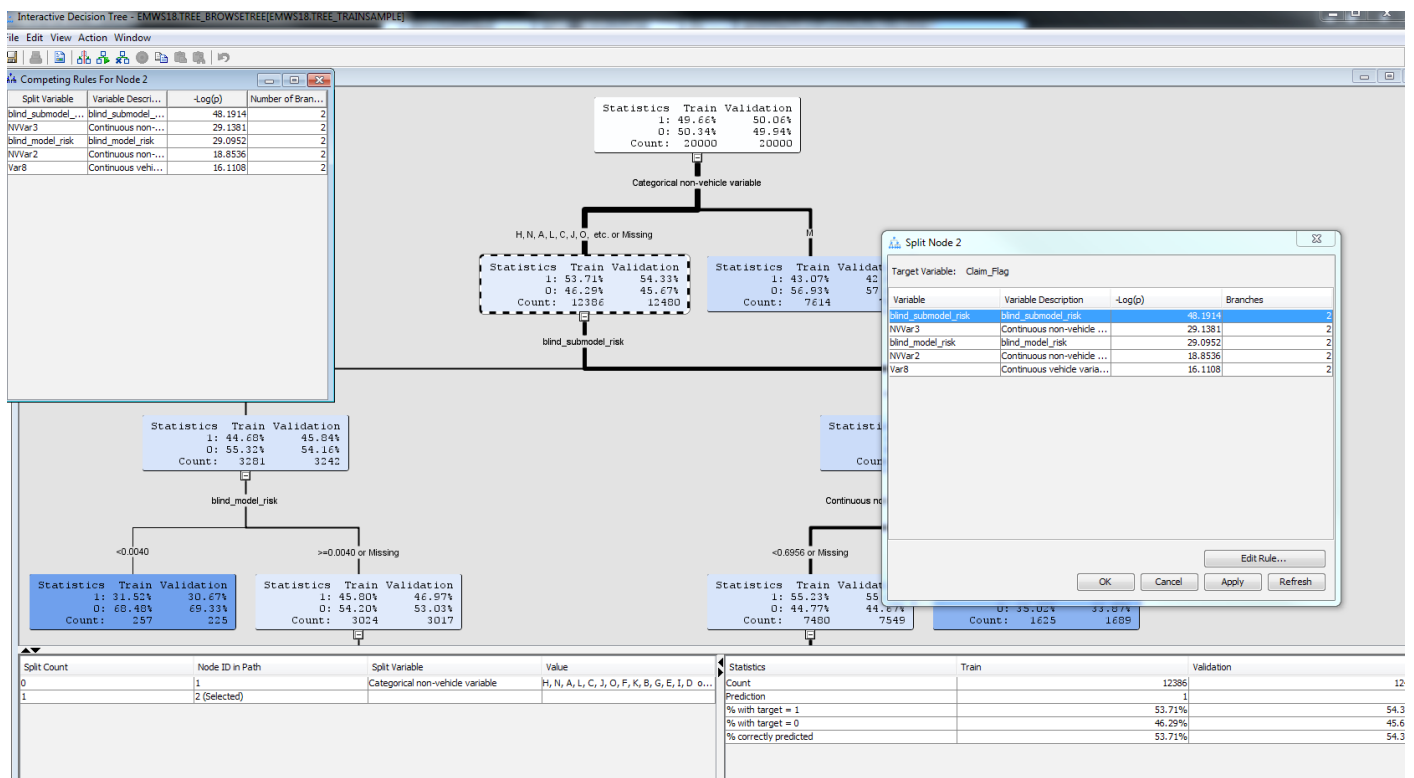


Figure 3: Decision trees are just one of the many modeling techniques included with SAS Enterprise Miner. They can be developed interactively or in batch mode. Numerous assessment plots help gauge overall tree stability.

## Using SAS® Factory Miner for an Automated Approach to Data Mining

As organizations apply more targeted analytics to their growing number of customer and business segments, there is a need to create even more predictive models at more granular levels. For example, instead of developing one model for the entire customer base, marketing departments want to create specific models for many customer segments. A retailer may want to develop cross-sell models for a large number of product categories. Or, a transport enterprise will want to build predictive maintenance models for different components of the vehicles it has in operation. And while this makes it necessary to create a lot more models, most analysts and data scientists don't have the luxury of more time.

With SAS Factory Miner, you get an interactive predictive modeling environment that makes it extremely easy to create, modify and assess hundreds, or even thousands, of models very quickly. With just a few clicks, you can access, modify and transform your data, choose which machine-learning techniques you want to apply and run the models in an automated model tournament environment to quickly identify the best performer for each segment. Modeling techniques included in SAS Factory Miner are:

- Bayesian networks.
- Decision trees.
- Gradient boosting.
- Neural networks.
- Random forests.
- Support vector machines.
- Generalized linear models.
- Linear regression.
- Logistic regression.

Users can easily identify modeling exceptions (segments where the automated approach does not generate models that meet acceptance criteria). The white-box design of SAS Factory Miner lets users easily modify predictive modeling pipelines and fine-tune parameters of pipeline components for better results where required. They can even create their own customized modeling pipelines for their favorite analytical projects, including data preparation, feature engineering and selection and learning algorithms, and share them with other users to create a repository of organizational best practices. This collaboration across the entire organization can help expand the analytics talent pool in your organization. The data scientist acts

### With SAS Factory Miner, you can:

- Boost discovery productivity.
- Automate model development.
- Explore new ideas faster.
- Collaborate with your analytics peers across your organization.
- Expand your analytics talent pool through automated self-service machine learning.
- Put large predictive model portfolios into production more efficiently and manage them with ease.

as the producer of organizational best-practice modeling pipelines for different projects and other users of the environment consume these best practices in a self-service fashion for optimal results.

And SAS Factory Miner does not stop with the identification of a champion model for each segment. Complete code is automatically created for the entire scoring pipeline (including data transformations) of each model for deployment in SAS or other environments, such as databases or Hadoop.

In addition, all model development and scoring assets can be registered to SAS Decision Manager, a centralized web-based environment for managing the life cycle and governance of your modeling assets from SAS or third-party providers, including open-source analytics.

The automation, ease of use, scalability and collaboration capabilities of SAS Factory Miner ramp up your predictive model-building power, increase the productivity of your analytics staff, enable collaboration across dispersed analytics teams, as well as expand your analytics talent pool through the democratization of machine-learning techniques.

## Scaling Your Discovery Process to Handle Big Data and Complex Problems

Big data and complex problems call for big analytics solutions. At SAS, we amp up your discovery power with distributed in-memory analytics. The idea is simple yet powerful. Break your data into smaller chunks and distribute the volume of the data and the complexity of the problem across your compute

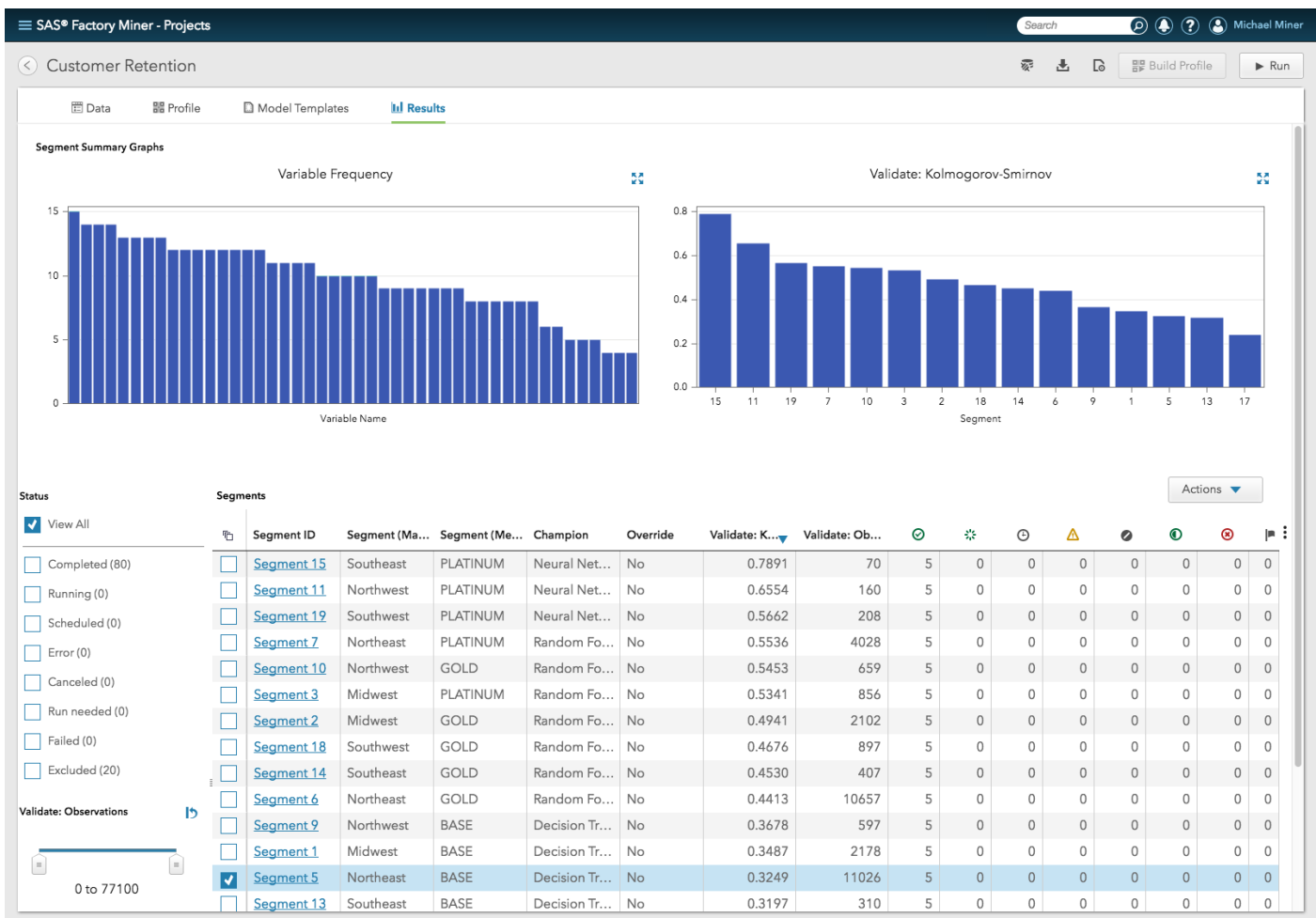


Figure 4: Customizable assessment techniques in SAS Factory Miner enable you to generate champion models for every segment in your data.

engines, whether it's on a single machine with a multitude of processing cores (CPUs) or a network of computers, such as a Hadoop cluster. The processing is done entirely in memory whenever possible, including the communication between the processing units (CPUs), which makes this process really fast.

SAS distributed, in-memory analytics processing takes advantage of a highly scalable and reliable analytics infrastructure – including database appliances like Pivotal Greenplum, Teradata, Oracle and SAP HANA – and commodity hardware using open source Hadoop or Hadoop Cloudera and Hortonworks distributions. For the users, nothing much changes. They can work from the same familiar interface for their data mining, predictive analytics and machine-learning projects, while SAS In-Memory Analytics takes care of the optimal workload distribution on the available system.

SAS High-Performance Data Mining lets you analyze large volumes of diverse data using a drag-and-drop interface and powerful descriptive, predictive and machine-learning methods. A variety of modeling techniques – including random forests, support vector machines, neural networks and clustering – are combined with data preparation, data exploration and scoring capabilities. Because you're able to build and run more models faster, you can ask more questions and bring new ideas into your data mining process. SAS High-Performance Text Mining lets you gain quick insights from large unstructured data collections involving millions of documents, emails, notes, report snippets, social media sources, etc. Support is included for parsing, entity extraction, automatic stemming and synonym detection, topic discovery and singular value decomposition (SVD). Text mining results can be used as inputs into high-performance data mining to improve your predictive modeling power.

## Integration Eases Model Deployment, Monitoring and Management

While this paper focuses on the data mining and analytical discovery process, you can't really end a conversation about data mining and machine learning for business applications without touching on what happens after the predictive models are built and the champion model chosen. So, what does happen? You move on to the deployment phase (see Figure 2).

After the champion model has been selected, it needs to be implemented into the right production environment. Organizations use predictive models in different ways. For example, they might be used to select customers for marketing campaigns by running a batch-scoring process and providing the selected customers as a list to marketing. An increasing number of organizations are looking into more integrated and automated processes to make the results of predictive models available for operational decision making. Rather than having the scoring process run in batch, they would like to have the model provide on-demand answers as part of a business application. Organizations may also want real-time answers from streaming data (e.g., for automated fraud detection or predictive maintenance).

Furthermore, business rules are being used in conjunction with analytical models to make decisions more flexible and agile. With SAS Decision Manager, business rules help define the actions based on specific conditions in business processes.

In the past, the deployment of a predictive model into the production environment has been manually performed by IT, often resulting in huge delays before the model could be used. With constantly changing market conditions and new data continuously arriving, it's possible for models to become obsolete before they are even deployed. With the seamless integration of the discovery and deployment phases of the analytical life cycle, SAS enables organizations to automate this process. SAS Decision Manager provides a streamlined interface to deploy models to execution environments in real time or batch without recoding the models for different environments. This maximizes the investment in the analytics through reuse of the analytical assets across environments and reduces risks by eliminating the need for manual recoding and subsequent revalidation: develop once, deploy many times.

Forward-thinking organizations are finding new ways to be more efficient and drive better automated decisions. SAS Decision Manager provides the features that organizations need for faster and easier model deployment into production situations.

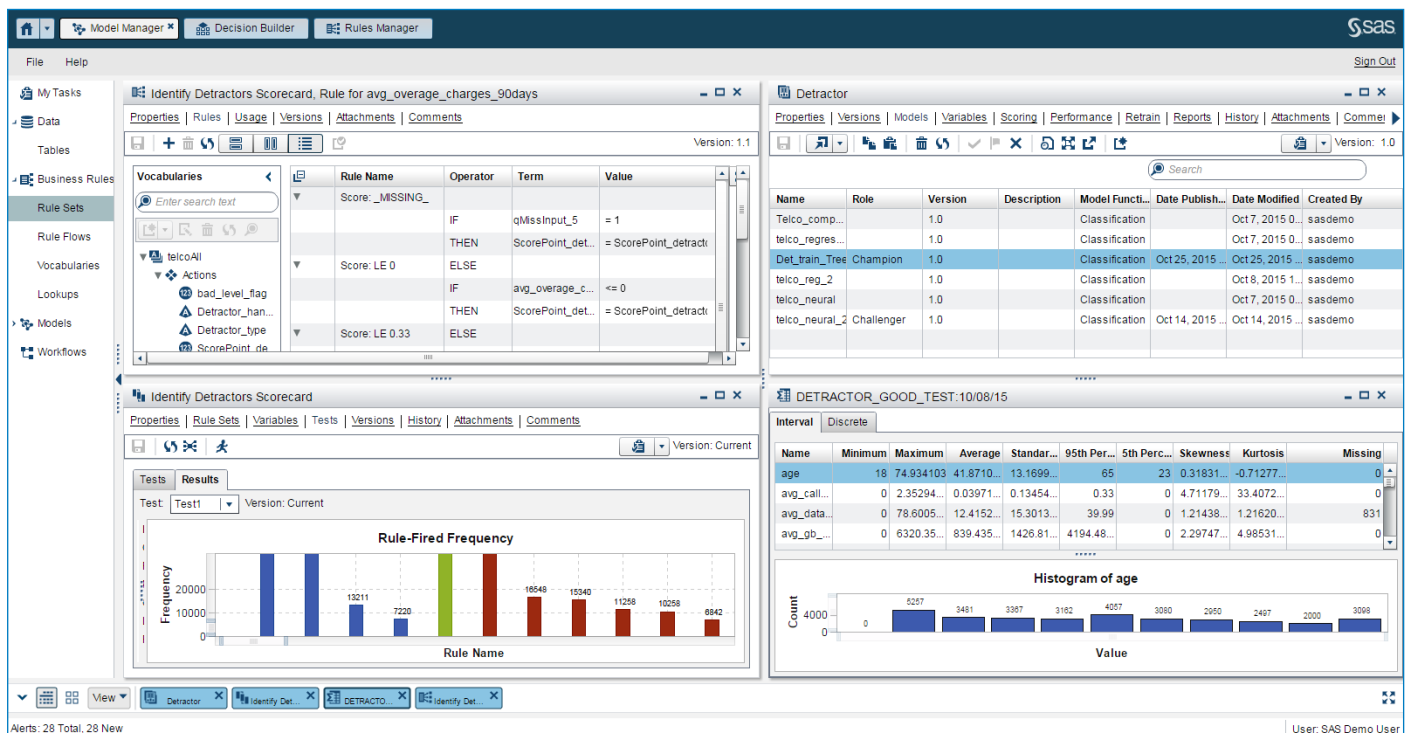


Figure 5: SAS Decision Manager helps expedite the model deployment process. It integrates model development automation with SAS Factory Miner and accelerates common manual tasks, like the definition of business rules and automatic generation of vocabularies.



## Conclusion

Today, more organizations are recognizing the value of predictive analytics results. And that's good because if you're collecting and storing data, you should be using it to gain insights that lead to competitive advantage.

This is especially true if your organization is paying people to create analytical models! But the trick has always been getting all the different pieces and parts moving together in order to extract the maximum value from all your data. SAS offers a complete analytical-lifecycle process that helps organizations go from data to decisions on a very large scale, in a very reliable manner.

It starts with data access and preparation (data volumes don't matter), moves through the process of data discovery and analytical modeling to produce predictive insights, and goes on to the deployment and management of results – all in an integrated environment.

While this paper introduced all phases of the analytical life cycle, its main focus was on the discovery portion. And at SAS, discovery means using predictive analytics to quickly and easily find new and reliable insights from data. With industry-recognized data mining software like SAS Enterprise Miner, the new SAS Factory Miner solution, in-memory technologies and enterprise model management capabilities, organizations are able to tackle any big data analytics problem.

- SAS Factory Miner provides an automated, web-based solution for building and retraining predictive models across multiple segments. It boosts productivity by enabling modelers to quickly and easily test many approaches simultaneously using machine learning and statistical algorithms.
- In situations where automated modeling doesn't work, SAS Enterprise Miner can be used to handcraft customized, strategic advanced predictive models.
- Distributed in-memory computing keeps processing moving at maximum speeds.
- SAS Decision Manager streamlines analytical model deployment – all from a single interface.

These solutions streamline the data discovery/data mining process, enabling you to create highly accurate predictive and descriptive models based on data analysis from across your enterprise.

## Learn More

Visit [sas.com/datamining](https://sas.com/datamining) to find out more about our data mining and data discovery solutions.

Join the [SAS Data Mining Community](#), where users and SAS employees share tips and other information.

For a complete overview of the entire analytical life cycle, read [Manage the Analytical Life Cycle for Continuous Innovation](#).

To learn more about the deployment phase, read [From Data to Decision: How SAS® Decision Manager Automates Operational Decisions](#).

To contact your local SAS office, please visit: [sas.com/offices](https://sas.com/offices)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2016, SAS Institute Inc. All rights reserved.  
104937\_S149733.0116

