

WHITEPAPER

DATA MINING WITH MICROSTRATEGY

USING THE MICROSTRATEGY BI PLATFORM TO DISTRIBUTE
DATA MINING AND PREDICTIVE ANALYTICS TO THE MASSES

DATA MINING WITH MICROSTRATEGY

USING THE MICROSTRATEGY BI PLATFORM TO DISTRIBUTE DATA MINING AND PREDICTIVE ANALYTICS TO THE MASSES

Executive Summary	5
1. Introduction	6
2. Overview of Data Mining	6
2.1 What is Data Mining?	6
2.2 What is the Data Mining Process?	7
2.3 Applications of Data Mining Integrated with Business Intelligence	8
2.4 Types of Data Mining Algorithms	9
2.5 Integrated Data Mining and Business Intelligence	10
2.6 Predictive Model Markup Language (PMML)	12
3. The MicroStrategy Business Intelligence Platform	12
4. Data Mining Services in the MicroStrategy BI Platform	14
4.1 MicroStrategy's Built-in Data Mining Functions	15
4.2 MicroStrategy's Data Mining Integration using PMML	16
Conclusion	18

EXECUTIVE SUMMARY

Data Mining is a broad term often used to describe the process of using database technology, modeling techniques, statistical analysis, and machine learning to analyze large amounts of data in an automated fashion to discover hidden patterns and predictive information in the data. By building highly complex and sophisticated statistical and mathematical models, organizations can gain new insight into their activities, such as:

- What are my sales forecasts for next year?
- Is this new customer a credit risk for my organization?
- Which two items sell well together?
- Which customers are most likely to respond to my next credit card promotion?

More and more, organizations are using data mining to make proactive knowledge-driven decisions, improving their organization's efficiency and effectiveness. Despite the power of data mining tools, growth and user adoption of these applications has been low. This can be attributed mainly to their specialized application, which until now, has remained in the realm of highly qualified statisticians. They also lack BI functionality, proactive information distribution and collaboration, robust security, easy self-service analysis, and are unable to scale to large user populations and data volumes.

The MicroStrategy BI platform delivers data mining to the masses through its Data Mining Services. It empowers all users to perform data mining by using metrics built with out-of-the-box predictive functions or imported data mining models from 3rd party data mining tools. Since Data Mining Services is fully integrated in the MicroStrategy BI platform, highly formatted and data rich data mining reports can be accessed and viewed through a wide variety of different interfaces, namely: Web, email, portal, Excel, etc. In addition, standard MicroStrategy BI functionality, such as slice-and-dicing data, ad-hoc report creation, drilling, pivoting, filtering and sorting, is available on these predictive reports, just as they are on any other report. With MicroStrategy, predictive analytics can now be distributed throughout the entire organization.

1 . INTRODUCTION

The purpose of this document is to provide users with a background of a few key data mining concepts, information about Data Mining Services in the MicroStrategy BI platform, and business scenarios to illustrate MicroStrategy's approach to data mining analysis. By the end of this document, users will have an understanding of the important concepts required to build highly sophisticated data mining and predictive analysis reports.

This document is intended for MicroStrategy customers and prospects who wish to:

- Learn how MicroStrategy allows decision makers to distribute data mining and predictive analytics across their entire organization,
- Understand how MicroStrategy integrates with existing data mining software, so they can leverage their existing data mining investment, and
- Perform data mining and predictive analytics without 3rd-party data mining software.

2 . OVERVIEW OF DATA MINING

2.1 What is Data Mining?

Data Mining, or knowledge discovery, is the process of examining large amounts of data in search of hidden patterns and predictive information in an automated manner. This information allows organizations to make better decisions. Data Mining uses database technology, modeling techniques, statistical analysis and machine learning to:

1. Find hidden patterns and make predictions which elude all but the most expert users.
2. Generate scoring or predictive models based on actual historical data.

Most Fortune 500 companies utilize highly sophisticated data mining applications to monitor and develop their business activities. Although each organization's data mining application serves a specific purpose, these applications all have something in common; they all leverage information about customers or business processes stored in data warehouses to reduce costs, improve the value of their customer relationships, and reduce the exposed risk of the organization.

Some real-world examples of data mining applications include:

- A telecommunications company saved over \$1 million in phone repairs by finding a trend in the type of repair problems and making a process shift to address this change.
- A financial institution developed a credit scoring model based on decision rules to reduce the risk of loan defaults.
- A leading insurance provider saved over \$100 million by identifying fraudulent claims. They were able to identify specific behaviors and trends in those that submitted fraudulent claims and apply these rules to identify suspicious activity in millions of claims.
- A state police department was able to automate the process of analyzing millions of telephone calls and internet records to identify suspicious activity.

2.2 What is the Data Mining Process?

How does data mining find information that business users and analysts did not already know? How does it find information about what is likely to happen next? This process is done through a technique called modeling and involves the following three steps:

1. Create a predictive model from a data sample.
2. Train the model against datasets with known results.
3. Apply the model to a new dataset with an unknown outcome.

In general, data mining software assists and automates the process of building and training highly sophisticated data mining models and applying these models to larger datasets. For example, suppose that a credit card company plans to develop a promotional campaign to recruit new customers, let's look at the three steps used to build a predictive model.

1. **Create a predictive model from a data sample** – A sample dataset of customers who have responded to past promotional campaigns is extracted from the data warehouse. This sample contains customer characteristics and trends that potentially can be used to predict “responsiveness”. For example:

- Where do they live?
- What gender are they?
- What age range do they fall in?
- What is their income range?
- What is their marital status?
- What is their level of education?
- What are their past purchases?
- Have they responded to past campaign?
- What is their credit history?

Advanced statistical and mathematical techniques are used to identify the significant characteristics and trends in predicting responsiveness, and a predictive model is created using these as inputs. Note that often only a small subset of the all characteristics and trends in the sample dataset are generally used in the model.

2. **Train the model against datasets with known results** – The new predictive model is applied to additional data samples based on historical campaign responses. This gives a good indication on how accurate the model is. It can then be further trained using these samples to improve its accuracy.
3. **Apply the model against a new dataset with an unknown outcome** – The predictive model is applied to the new customer/prospect database. The end result will yield those customers that have a high probability of responding to the marketing promotion.

2.3 Applications of Data Mining Integrated with Business Intelligence

To understand the power of data mining and how business intelligence allows this information to be distributed to all relevant decision makers, it is helpful to look at various different use cases and business examples.

- **Market Basket Analysis** – This effective data mining modeling technique is used to determine items that are frequently sold together. Using association rules, a nationwide grocery store identified hidden patterns in buying behavior that had been previously overlooked. The implication of these findings suggested that store managers should place items that are often purchased together in key strategic locations across the store to promote the sales of these items.
- **Fraud Detection through Purchase Sequences** – A major credit card company introduced a new offering to protect their customers against fraud. They used a sequence association model to detect fraudulent purchases. By analyzing historical data, they noticed that when a transaction for a gas purchase was followed by transactions for expensive luxury items, there was a high probability that these purchases were fraudulent. The new product offering used a series of these rules that identified potential fraudulent activity which protected their customers against unauthorized purchases.
- **Campaign Management** – A mail-order retailer wanted to improve the effectiveness of its direct mail marketing campaigns, with the goals of reducing costs and increasing the percent of positive responses. The retailer knew that it is too costly to send direct mail to all of its customers. Using a neural network model, they analyzed all of the factors that affect their customer's propensity to respond. The model included many variables, such as past purchase history, purchase frequency, customer age, gender, marital status, location, etc. and it was trained on a number of historical mailing campaigns. The model was then applied to the full list of customers and the probability of them responding to the campaign was predicted. Customers marked as most likely to respond were targeted in the new campaign.
- **Instant Credit Scoring** – A commercial bank wanted to automate the process of approving loan applications to save costs. Through a series of if-then rules using a decision tree, a credit score was generated for each new loan application which helped to identify whether it should be approved. This application decreased their costs by employing fewer customer service representatives and increased customer service ratings by allowing customers to know instantly whether their loan had been approved.
- **New Restaurant Locations** – A nationwide fast food restaurant chain used data mining models to determine the best place to establish new restaurants. They grouped together all of the variables that are likely to influence the sales of a new restaurant. These included variables like: population size and demographics, competition, distance from other franchises, etc. Using a regression model, they were able to input a prospective restaurant location and estimate the potential growth and profitability of this location. They compared the outputs of various prospective locations to identify which had the highest profitability potential. Prior to using this data mining model, the company relied on educated guesses backed by the demographic data of the location.
- **Television Audience Share Prediction** – A nationwide television programming station needed to predict the audience share of a new TV program which was scheduled for broadcast at a particular time. With years of historical data containing audience share for each program shown in each time slot, a neural network model based on a large number of variables was developed to predict the audience share. These variables included: the characteristics of the new program, such as genre, time of showing, target audience, cast, etc., the preceding and following programs with their characteristic information, other programs shown at the same time and the audience share, time of year, major public and sporting events, weather, etc. The model was able to predict audience share accurately which resulted in better sales opportunities for advertising slots.

- **Online Sales Improvement** – Online merchants rely on cross-selling and up-selling to increase their revenues. By relying on historical sales and user ratings of specific items, buyers are provided a choice of “similar” products when browsing specific items in the store. A nearest neighbor model generates “similarity” metrics which browse the product data warehouse for products nearest to a selected item, enticing the buyer to purchase additional items.
- **Failure Rates Prediction** – In an effort to reduce repair costs, an automobile manufacturer needed to be able to predict the failure rates on various automotive parts. Using a regression model, the manufacturer was able to identify and measure all of the variables that are likely to affect part failures. This enabled them to determine if a part was likely to fail and when it was likely to happen. By identifying defective parts at an earlier stage in a car’s lifecycle, warranty repairs and costs were drastically reduced and satisfaction ratings improved among the automobile owners.

2.4 Types of Data Mining Algorithms

There are a wide range of different data mining algorithms in use today. Each algorithm can be used in a wide variety of applications and it is up to a data mining specialist to choose which algorithm, or combination of algorithms, will be the best fit to develop, test and train a predictive model.

- **Regression** is a powerful and commonly used algorithm that evaluates the relationship of one variable, the dependent variable, with one or more other variables, called independent variables. By measuring exactly how large and significant each independent variable has historically been in its relation to the dependent variable, the future value of the dependent variable can be estimated. Regression models are widely used in applications, such as seasonal forecasting, quality assurance and credit risk analysis.
- A **Neural Network** is a sophisticated pattern detection algorithm that uses machine learning techniques to generate predictions. This technique models itself after the process of cognitive learning and the neurological functions of the brain capable of predicting new observations from other known observations. Neural networks are very powerful, complex, and accurate predictive models that are used in detecting fraudulent behavior, in predicting the movement of stocks and currencies, and in improving the response rates of direct marketing campaigns.
- A **Decision Tree** is a tree-shaped graphical predictive algorithm that represents alternative sequential decisions and the possible outcomes for each decision. This algorithm provides alternative actions that are available to the decision maker, the probabilistic events that follow from and affect these actions, and the outcomes that are associated with each possible scenario of actions and consequences. Their applications range from credit card scoring to time series predictions of exchange rates.
- **Clustering / Segmentation** is the process of grouping items together to form categories. You might look at a large collection of shopping baskets and discover that they are clustered corresponding to health food buyers, convenience food buyers, luxury food buyers, and so on. Once these characteristics have been grouped together, they can be used to find other customers with similar characteristics. This algorithm is used to create groups for applications, such as customers for marketing campaigns, rate groups for insurance products, and crime statistics groups for law enforcement.
- **Association Rules** detects related items in a dataset. Association analysis identifies and groups together similar records that would otherwise go unnoticed by a casual observer. This type of analysis is often used for market basket analysis to find popular bundles of products that are related by transaction, such as low-end digital cameras being associated with smaller capacity memory sticks to store the digital images.

- **Sequence Association** detects causality and association between time-ordered events, although the associated events may be spread far apart in time and may seem unrelated. Tracking specific time-ordered records and linking these records to a specific outcome allows companies to predict a possible outcome based on a few occurring events. A sequence model can be used to reduce the number of clicks customers have to make when navigating a company's website.
- **Nearest Neighbor** is quite similar to clustering, but it will only look at others records in the dataset that are "nearest" to a chosen unclassified record based on a "similarity" measure. Records that are "near" to each other tend to have similar predictive values as well. Thus, if you know the prediction value of one of the records, you can predict its nearest neighbor. This algorithm works similar to the way that people think – by detecting closely matching examples. Nearest Neighbor applications are often used in retail and life sciences applications.

2.5 Types of Data Mining Algorithms

Data mining is not very wide spread even though many organizations use some sort of data mining tool for their predictive analysis. This can be attributed to their specialized application, which until now, has remained in the realm of highly qualified statisticians. The non-integrated nature of data mining tools and the difficulty of deploying data mining models to the general user population have caused user adoption of these applications to be extremely low. Data mining tools lack standard BI functionality, robust security, flexible information distribution and collaboration and easy self-service analysis, and their inability to scale to large user populations and data volumes have further hindered the growth and deployment of these applications.

IT departments are now starting to realize the tremendous value that an integrated data mining and BI application can bring their organization. In the past, there was a clear separation between the two applications, but the desire to share information and analytics from these applications is growing. Data that was once in the hands of data mining experts is now required by all of the relevant decision makers, managers, and stakeholders to improve the accuracy of their decisions. The integration of data mining's predictive analysis with BI applications in the MicroStrategy BI platform provides the ability to:

- Create flexible, organized and highly formatted predictive reports for easiest possible user consumption and professional presentation.
- Deliver individualized messages and predictive reports to very large populations through a single 'service definition' – based on event triggers or schedules.
- Provide large user populations with ad-hoc query and ad-hoc analysis on predictive data of an entire database without requiring knowledge of SQL, table structures, or predictive models.
- Slice and dice predictive information within a limited analytical domain making it simple and safe for causal users.
- Implement even the strictest security scheme to users within and outside the organization.
- Apply data mining models against terabytes of data.

Scoring and Integration Options

Once a data mining model has been created and tested, it is applied to a new dataset. The process of computing this predictive model to produce a final outcome is called "scoring". There are three common methods of scoring

data. With each approach, a different application is responsible for the model scoring. While each approach has its advantages and disadvantages, it is up to the analyst and IT administrator to determine which approach is most suitable for their environment and their use cases.

A. Data Mining Tool Scoring

With this approach, the data mining tool does the scoring and the results are written into a database table. Once stored in the database, a BI application can easily access this new data and report designers can build predictive reports off this information to distribute to the relevant business users.

Advantages	Disadvantages
<ul style="list-style-type: none"> The Data Mining tool's scoring engine does the scoring leveraging the model's complexity. The scoring engine also takes into account performance considerations specific to the data mining tool. 	<ul style="list-style-type: none"> Requires database space and administrative support to create, store, and maintain these database tables. Updating new models and scoring requires the database administrator's assistance and takes time to implement.
<ul style="list-style-type: none"> The BI application only needs to read and display the information from the database, without having to perform complex calculations on the fly. 	<ul style="list-style-type: none"> New database records are not be scored. The scoring needs to be done on a frequent basis to leverage the new data.
<ul style="list-style-type: none"> BI application can take advantage of this predictive information by mapping their application to these new scoring tables. 	<ul style="list-style-type: none"> Difficulty scaling to large datasets

B. Database Scoring

This approach involves leveraging the database's existing data mining capabilities for scoring. Database tools, like IBM® Intelligent Miner and Teradata® Warehouse Miner, provide algorithms for analysts to build their own data mining models.

Advantages	Disadvantages
<ul style="list-style-type: none"> Can scale to very large datasets. Databases are designed to process large amounts of data in an efficient manner. 	<ul style="list-style-type: none"> Requires database space and administrative support to create, store, and maintain these database tables.
<ul style="list-style-type: none"> Scores can be calculated "on the fly" as new records are added to the database. 	<ul style="list-style-type: none"> Requires specific application knowledge of the database data mining tool.
<ul style="list-style-type: none"> Models are stored in a central location, database server, along with all of the data. 	<ul style="list-style-type: none"> Not all database vendors offer these capabilities. Those vendors that do support scoring may only support a limited number of algorithms.

C. Business Intelligence Application Scoring

The final approach is to have the BI application do all of the scoring. MicroStrategy offers native data mining capabilities allowing users to either build their own data mining algorithms or import them from 3rd party data mining tool directly into the BI application. Once imported, reports are built using these models and the models will be scored each time the report is executed.

Advantages	Disadvantages
<ul style="list-style-type: none">• Predictive metrics and reports existing in the application's metadata and can be used like any other report or metric.	<ul style="list-style-type: none">• Difficulty scoring large number of records. When dealing with many records, it is best to score on the database.
<ul style="list-style-type: none">• Scores are always calculated "on the fly" as new records are added to the database.	<ul style="list-style-type: none">• Predictive models need to be imported into the BI application.
<ul style="list-style-type: none">• Does not impact the database. No need for database administrative support to design or maintain database tables.	

2.6 Predictive Model Markup Language (PMML)

A key innovation which enables the integration of data mining models with other applications is Predictive Model Markup Language, or PMML. PMML is an XML standard that represents data mining models developed and trained by data mining tools. This industry standard was developed by the Data Mining Group (DMG), an independent group of over two dozen leading technology companies, including MicroStrategy. PMML supports a number of different data mining algorithms, including Regression, Neural Networks, Clustering, Decision Trees and Association, and incorporates data transformation and descriptive statistics. PMML is generated by nearly all data mining tools from companies such as SAS®, SPSS®, Microsoft®, Oracle®, Teradata®, IBM®, Quadstone, and others. MicroStrategy is the first Business Intelligence platform to support the PMML standard and, by allowing predictive metrics to be accessible by all users in the enterprise, **data mining for the masses** is possible.

3. THE MICROSTRATEGY BUSINESS INTELLIGENCE PLATFORM

The MicroStrategy BI Platform is the industry's only industrial-strength business intelligence architecture. MicroStrategy's high performance BI platform is accessed through a zero-footprint unified Web interface that provides all five styles of business intelligence and extends enterprise reporting and analysis to the masses.

All Five Styles of Business Intelligence

With MicroStrategy, companies have the most powerful and comprehensive business intelligence capabilities on one single platform. This means that they all share common infrastructure, common metadata definitions, common prompting, common scheduling, common caching, common security, etc, and that everything is re-used across each of the styles of BI, accelerating the development process and minimizing the number of servers required. Even

more important is the fact that all five styles of BI are presented to the user from a single integrated architecture through a unified Web interface. The five styles of BI are:

- Scorecards and Dashboards
- Reporting
- OLAP Analysis
- Predictive Analysis
- Alerting and Proactive Notification

Users have no knowledge that they are using different BI styles. Within a single look-and-feel, business people are finding reports, running reports, answering prompts, manipulating the results, and finally saving, printing, or exporting their work.

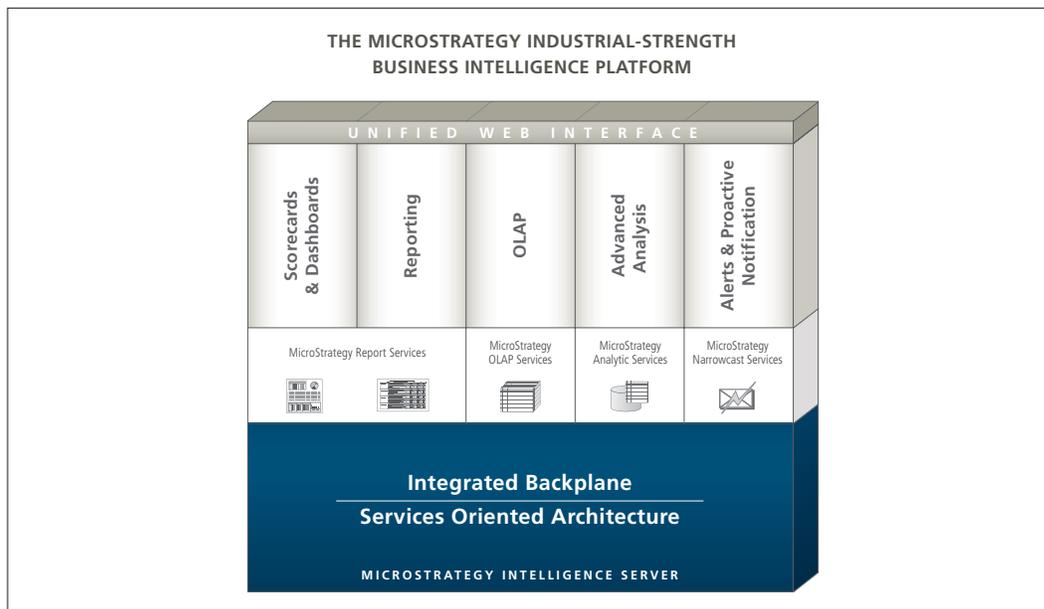


Figure 1: The evolution of BI applications has resulted in 5 Styles of BI, historically delivered with distinct BI technologies. Only MicroStrategy provides all 5 styles of BI with a unified Web user interface and an integrated backplane.

User Scalability

The MicroStrategy platform offers unparalleled user scalability – perfect for companies with hundreds of users and growing, and perfect for companies with hundreds of thousands of users internal and external to the enterprise. Deployment to all users worldwide is simple with MicroStrategy’s zero-footprint, pure Web interface. With MicroStrategy, all users have immediate and secure access to all reporting and analysis applications from any Web browser.

Data Scalability

The MicroStrategy platform provides unlimited data scalability, allowing companies to report on and analyze all enterprise data as it amasses into terabytes and beyond. MicroStrategy’s relational OLAP (ROLAP) architecture combined with its Intelligent Cube™ technology can handle any size database while delivering high performance. With MicroStrategy, companies can carry out comprehensive enterprise reporting and analyses down to the transaction level of detail.

Enterprise-Caliber IT

With features such as optimizations for very large databases, 24x7 uptime, a multi-pass SQL engine and an advanced prompting engine, MicroStrategy's architecture is renowned among IT administrators. MicroStrategy's integrated backplane is a centralized foundation for common metadata, prompting, scheduling, administration, privileges and bullet-proof security. MicroStrategy's scalable, server-centric architecture provides for the highest report throughput and most fault-tolerant operations. MicroStrategy gives companies the entire range of business intelligence functionality and the capabilities to deliver any or all of that functionality to an enterprise population of users, with complete centralized control and economies of administration.

4. DATA MINING SERVICES IN THE MICROSTRATEGY BI PLATFORM

Data Mining Services delivers data mining predictive models to all users of the MicroStrategy BI platform and is fully integrated with the platform. All decision makers can view, analyze, slice-and-dice, build, and distribute predictive reports through a zero-footprint Web interface, all with the highest level of security. Data Mining Services goes beyond traditional data mining workflow and incorporates data mining and predictive analysis into a single BI platform.

Data Mining Services provides predictive analysis in two distinct ways. With both approaches, MicroStrategy does the scoring "on the fly" without the need for database administrative support.

1. Users can use any one of MicroStrategy's 250 analytical functions to build out-of-the-box data mining reports with MicroStrategy. MicroStrategy provides all of the resources necessary to create these commonly-used data mining reports without the need for data mining tools. Organizations that do not have data mining applications can use MicroStrategy pre-built functions, most notably regression and market basket analysis, to perform common data mining analysis. Data Mining Services creates its own PMML from these predictive metrics that can be exported to a 3rd party data mining tool for further analysis.
2. MicroStrategy complies with PMML, the industry standard XML-based language for data mining models. Data mining models from a 3rd party data mining tool can be imported directly into MicroStrategy and predictive metrics are automatically created in the metadata repository. Report designers and end users alike can build any report using these predictive metrics. By complying with the industry standard PMML, organizations can leverage their existing data mining applications to build domain specific and highly sophisticated models while still relying on MicroStrategy for its enterprise-caliber BI reporting and analysis.

Benefits of Integrating Data Mining with MicroStrategy's Industrial-Strength BI Platform

- Allows businesses to view predictive reports through a wide variety of different user interfaces, namely: Web, email, portal, Excel, etc.
- Delivers all 5 styles of BI through a single unified backplane and a single unified Web interface.
- Provides unlimited scalability, allowing organizations to monitor, report, and analyze predictive information against all enterprise data and deliver these reports to thousands of users.
- Helps businesses better align people to organizational goals by providing predictive information through easy-to-use products.
- Notifies business users of new predictive information through alerting and proactive report distribution.

- Creates highly formatted and data-rich predictive scorecards, dashboards, and managed metric reports for all corporate performance management needs.
- Empowers analysts to perform further predictive analysis, such as slice-and-dicing data, ad-hoc report creation, drilling, pivoting, and sorting on predictive reports.
- Provides the most robust security for intranet and extranet applications.

4.1 MicroStrategy's Built-in Data Mining Functions

MicroStrategy contains a library with over 250 basic, OLAP, mathematical, financial and statistical functions that can be used to create business metrics and key performance indicators. Other features such as set analysis, transformations, and collaborative analytics can be used in conjunction with the function library to provide data mining analysis without the need for a 3rd-party data mining tool.

An example of an out-of-the-box MicroStrategy data mining function is the multiple regression model, which supports linear, logarithmic and exponential regression between several independent variables and a dependent variable. This allows business users to create and view richly formatted reports with forecasted numbers.

The PMML for this new predictive metric is automatically generated by MicroStrategy and can be exported and used in other data mining tools for further tuning and analysis after which it can be imported back into MicroStrategy.

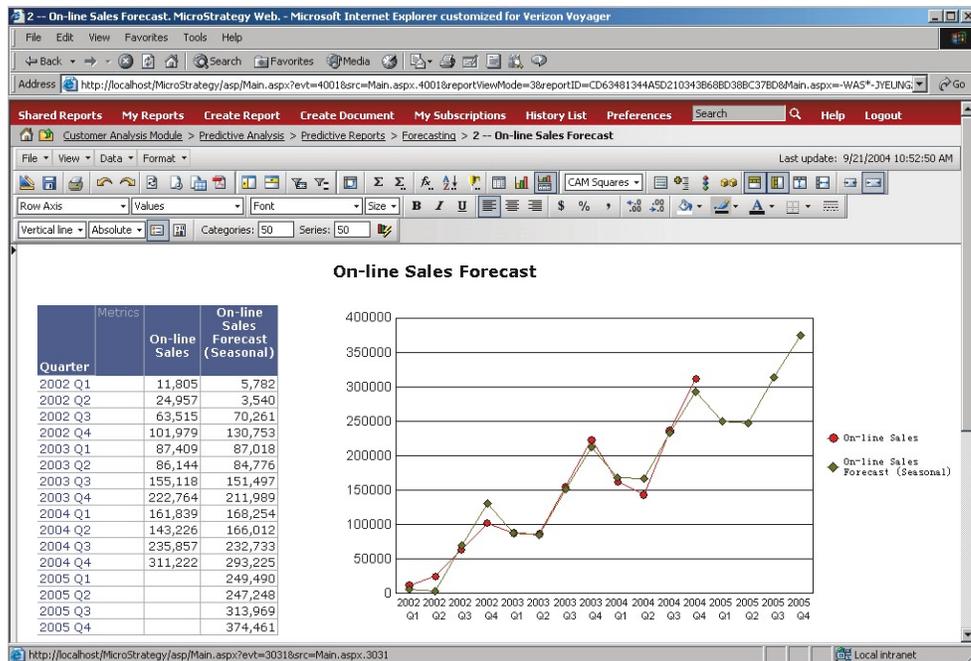


Figure 2: Business users can create and view predictive reports, just as they would with any other MicroStrategy report.

4.2 MicroStrategy's Data Mining Integration using PMML

Data Mining Services allows users to import PMML from 3rd party data mining tools, which can then be used to create predictive reports.

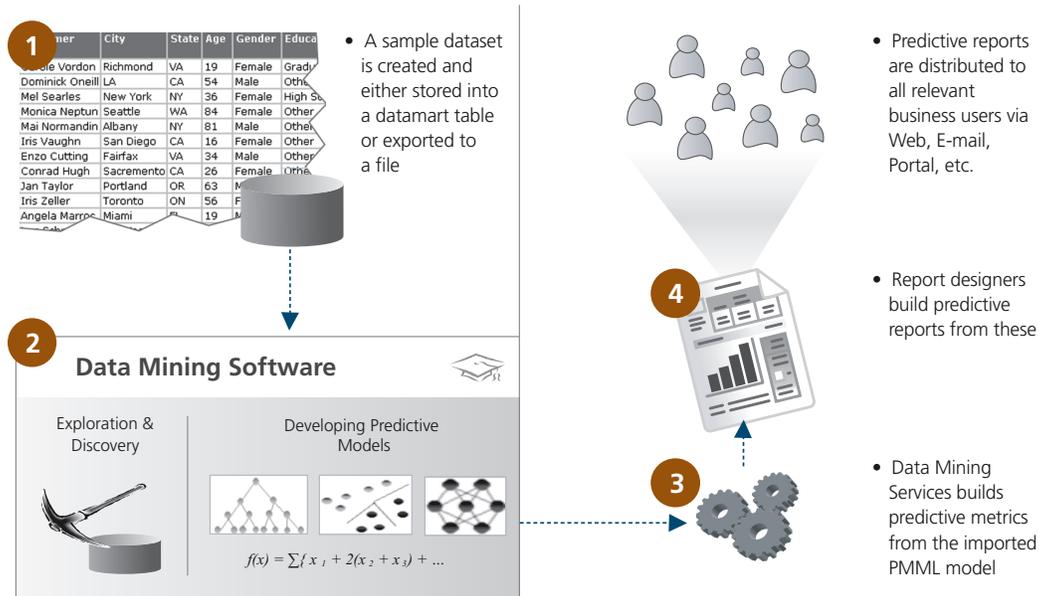


Figure 3: Workflow of Creating Predictive Reports using MicroStrategy and 3rd party Data Mining Tools.

1. Create a Sample Dataset and Export it to a 3rd party Data Mining tool

The first step is to create a sample dataset and store the dataset in a datamart table. Using MicroStrategy, a user creates a grid report by placing all of their desired metadata objects, such as attributes, metrics, filters, on this report. The user can run this report and drill down or across to other hierarchies to get their desired data sample. Next, the dataset report is exported in various formats, such as text files or Microsoft Excel, or saved to a database table using a dynamic datamart. Dynamic datamarts are a preferred approach when dealing with large datasets, whereas exporting to Excel or a text file may be more convenient for smaller datasets.

2. Using 3rd party Data Mining Tools to Develop Predictive Models

The user imports the dataset into the 3rd party data mining tool through an ODBC connection to the datamart table in a database or by simply uploading a file. A domain expert data mining analyst or statistician analyzes the dataset in search of predictive information and builds a predictive model from this dataset. Finally, the predictive model is exported to a file as PMML.

3. Data Mining Services import tool interprets the PMML and builds predictive metrics

Using MicroStrategy, the user imports the predictive model by uploading the PMML file. Once imported, Data Mining Services automatically creates a predictive function off of this PMML model. This new predictive function is stored in the MicroStrategy metadata and is treated as any "normal" metric. By complying with PMML, MicroStrategy works with all industry-leading data mining software vendors.

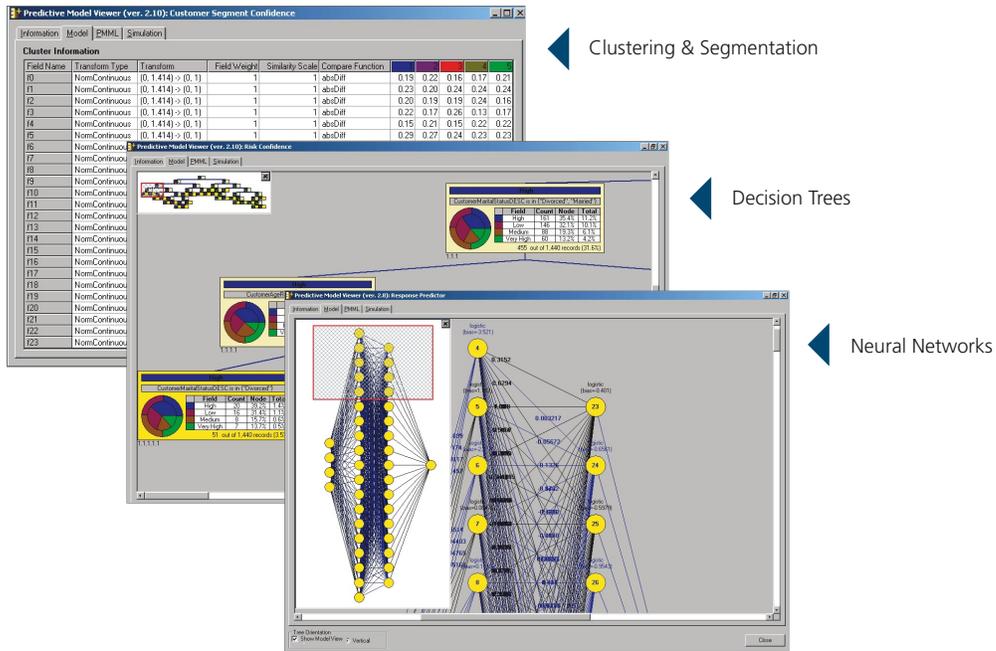


Figure 4: Data Mining Services provides users with a graphical representation of their predictive models. Shown above are a few examples of predictive models.

4. Building and Distributing Predictive Reports across your entire organization

Report designers can link this new predictive report to their original dataset and distribute this information to large groups of users. Predictive reports are viewed in Web, Office, Windows, and email/printer/file, just like any other MicroStrategy report. All manipulations are available, such as sort, filter, derived metrics, thresholds, etc.

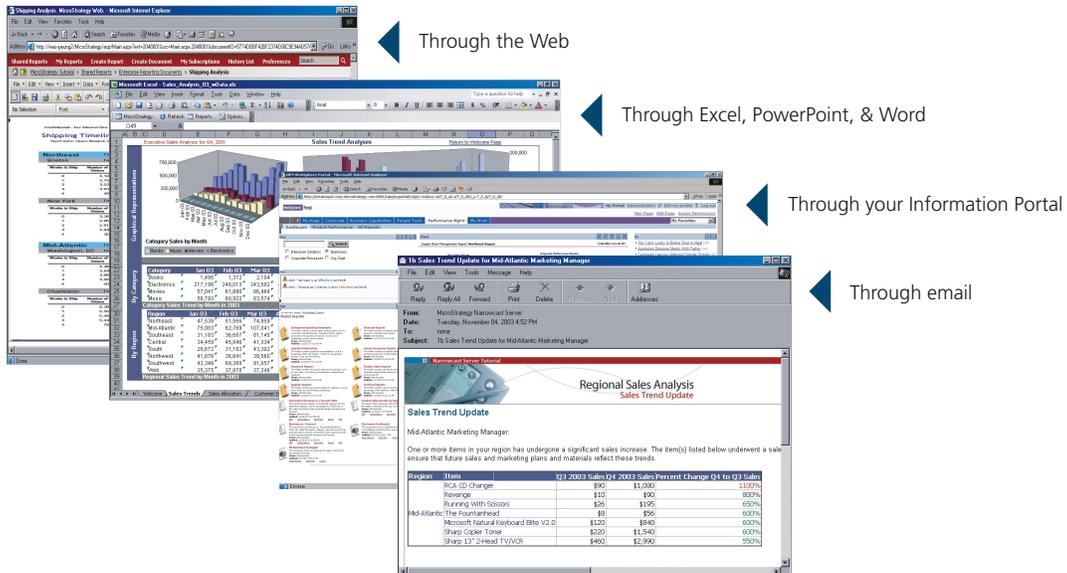


Figure 5: Since Data Mining Services is fully integrated into the MicroStrategy BI platform, MicroStrategy can deliver predictive reports through any interface of their choosing. Reports can be viewed through the Web, Excel and other Microsoft Office applications, enterprise information portals, email, and other wireless devices.

CONCLUSION

Many organizations have adopted data mining and predictive analytics applications to make proactive knowledge-driven decisions, improving their organization's efficiency and effectiveness. Despite the power of these data mining tools, the growth and user adoption of these tools has been slow due to their lack of BI functionality, proactive information distribution and collaboration, robust security, and easy self-service analysis, and their inability to scale to large user populations and data volumes.

The industry leading MicroStrategy BI platform is the first to deliver data mining and predictive analytics to all users through a fully integrated enterprise caliber BI system. Using Data Mining Services, business users, report designers, and analysts alike can view and build predictive reports using MicroStrategy and distribute these reports to all relevant decision makers and stakeholders.

COPYRIGHT NOTICE Copyright ©2005 MicroStrategy Incorporated, 1861 International Drive, McLean, Virginia, 22102 U.S.A. All rights reserved.

The material presented herein is based on information we consider reliable. We do not represent that it is accurate and complete. No person should consider our distribution of this material as making any representation or warranty with respect to such material and should not rely upon it as such.

This product is patented. One or more of the following patents may apply to the product sold herein: US Patent Nos. 6,154,766, 6,173,310, 6,260,050, 6,263,051, 6,269,393, 6,279,033, 6,501,832, 6,567,796, 6,587,547, 6,606,596, 6,658,093, 6,658,432, 6,662,195, 6,671,715, 6,691,100, 6,694,316, 6,697,808, 6,704,723, 6,707,889, 6,741,980, 6,765,997, 6,768,788, 6,772,137, 6,788,768, 6,792,086, 6,798,867, 6,801,910, 6,820,073, 6,829,334 and 6,836,537. Other patent applications are pending.

TRADEMARKS: MICROSTRATEGY, STRATEGY.COM, INTELLIGENT E-BUSINESS, MICROSTRATEGY WEB BUSINESS ANALYZER, MICROSTRATEGY ECRM 7, MICROSTRATEGY WEB, MICROSTRATEGY TELECASTER, MICROSTRATEGY AGENT, MICROSTRATEGY WORLD, MICROSTRATEGY INTELLIGENCE SERVER, MICROSTRATEGY BROADCASTER, MICROSTRATEGY ARCHITECT, MICROSTRATEGY ADMINISTRATOR, MICROSTRATEGY INFOCENTER, MICROSTRATEGY SDK, MICROSTRATEGY TRANSACTOR, MICROSTRATEGY 7, MICROSTRATEGY 7i, MICROSTRATEGY 8, BEST IN BUSINESS INTELLIGENCE, THE SCALABLE BUSINESS INTELLIGENCE PLATFORM BUILT FOR THE INTERNET, THE INTEGRATED BUSINESS INTELLIGENCE PLATFORM BUILT FOR THE ENTERPRISE, MICROSTRATEGY BUSINESS INTELLIGENCE PLATFORM, THE PLATFORM FOR INTELLIGENT E-BUSINESS, THE POWER OF INTELLIGENT E-BUSINESS, THE FOUNDATION FOR INTELLIGENT E-BUSINESS, INTELLIGENCE THROUGH EVERY PHONE, PERSONALIZED INTELLIGENCE PORTAL, INTELLIGENCE TO EVERY DECISION MAKER, APPLICATION DEVELOPMENT AND SOPHISTICATED ANALYSIS, CENTRALIZED APPLICATION MANAGEMENT, RAPID APPLICATION DEVELOPMENT, PERSONAL INTELLIGENCE NETWORK, MICROSTRATEGY 6, MICROSTRATEGY CONSULTING, MICROSTRATEGY EDUCATION, MICROSTRATEGY SUPPORT, DSS BROADCASTER, ESTRATEGY, ETELECASTER, EBROADCASTER, THE INTELLIGENCE COMPANY, THE E-BUSINESS INTELLIGENCE PLATFORM, CHANGING THE WAY GOVERNMENT LOOKS AT INFORMATION, DSS TELECASTER, THE POWER OF INTELLIGENT E BUSINESS, DSS BROADCASTER SERVER, DSS SUBSCRIBER, INFORMATION LIKE WATER, DSS AGENT, DSS ARCHITECT, DSS WEB, DSS SERVER, DSS OFFICE, QUERY TONE, QUICKSTRIKE, INSIGHT IS EVERYTHING, ALARM, ALARM.COM, ANGEL, ANGEL.COM, TELEPATH INTELLIGENCE, TELEPATH INTELLIGENCE, TELEPATH, ECASER, ALERT.COM AND IWAPU ARE EITHER TRADEMARKS OR REGISTERED TRADEMARKS OF MICROSTRATEGY INCORPORATED IN THE UNITED STATES AND CERTAIN COUNTRIES. OTHER PRODUCT AND COMPANY NAMES MENTIONED HEREIN MAY BE THE TRADEMARKS OF THEIR RESPECTIVE OWNERS.

