



Patricia Seybold Group

Trusted Advisors to Customer-Centric Executives

KXEN Analytical Framework 3.0

Data Mining Made Much Easier

By Mitchell I. Kramer

Sr. Vice President and Sr. Consultant, Patricia Seybold Group



KXEN Analytical Framework 3.0

Data Mining Made Much Easier

By Mitchell I. Kramer, Sr. Vice President and Sr. Consultant, Patricia Seybold Group April 22, 2004

NETTING IT OUT

Customer relationship analytic applications are among the most important elements of the analysis side of CRM. These applications help you understand your customer relationships. They give you the ammunition you need to take actions (on the operational side of CRM) that strengthen customer relationships, improve loyalty and satisfaction, and increase profitability.

The KXEN (Just say the letters: kay, ex, ee, en) Analytic Framework is a customer relationship analytic application (although it has broader application for any data mining application, too) made up of seven product modules. These modules implement, automate, and abstract the data mining process of preparing data, building and testing predictive models, and executing models to score your customer data.

The KXEN Analytic Framework has three key strengths: It's easy to learn and easy to use; its approach to data mining is highly automated and fast; and its architecture facilitates the integration of data mining capabilities into many types of business intelligence and even operational applications. It also has three main limitations: It's difficult to decide which predictive modeling approach to use for a given problem; its data preparation facilities, while automated, are not complete; and its integration facilities, while quite flexible, can be complex.

If you've been resisting data mining because you thought it was too complex, if OLAP isn't giving you consistent answers on a timely basis, or if you've been struggling to continue to

improve your customer relationships after an initial big hit, then KXEN Analytic Framework might be the analytic toolset for you.

NEEDED: EASY TO USE, HIGH-PERFORMANCE DATA MINING FOR CUSTOMER DATA

We've heard from you and we've long been hearing from data mining suppliers that most of you don't use data mining technologies. Cost, time, and, especially, complexity have been the obstacles in the way of broader adoption. It shouldn't be that way. Data mining is a technology that should be embraced. It's technology that has the potential to deliver the insight that you need to predict your customers' future behavior—behavior in terms of their likelihood to accept the offers of your campaigns and behavior in terms of their potential to desert your organization as customers.

KXEN Analytic Framework Removes the Obstacles

The KXEN Analytic Framework (KAF) can remove the obstacles that keep you from adding data mining to your analytic application toolkit. It's relatively low in cost: a hundred thousand dollars or so for a license. Because it's easy to learn and easy to use, you won't have to bear the addition of skilled and experienced data mining professionals needed for conventional data mining workbenches. Because it's fast, your data mining process can be completed in hours or a few days, not weeks or months. (We've seen it run.) Basically, you can purchase, implement, and use KXEN Analytic Framework for the same resources as for other types of analytic applications.

Have we got your attention? If so, then read on. It's about time you got serious about data mining.

UNDERSTANDING AND IMPROVING YOUR CUSTOMER RELATIONSHIPS

You use customer relationship analytic applications in order to understand the relationships that you have with your customers and to provide insight that can help you to improve these relationships. Customer relationship analytic applications can be critical to your business because they get directly to your bottom line. Improved customer relationships result in more satisfied, more loyal, and more profitable customers.

But understanding your customer relationships is not easy. Customer relationships cover a lot of ground. Your customers' behavior and transactions touch all of your channels, all of your customer-facing applications, and all the data, back-office, and partner applications that support them. Getting a handle on your customer relationships can take some work. Customer relationship analytic applications simplify, or even eliminate, some of that work.

Customer relationship analytic applications are tools that can help you accomplish the following:

- Identify your best customers (and your worst ones) by their loyalty, revenue, profitability, or any metric that you use
- Identify behavior and transactions that indicate which of your customers are at risk to desert you
- Group your customers to make your marketing campaigns and your service offerings more effective and more efficient
- Make offers that your customers will accept and that, at the same time, will increase your revenue or reduce your cost
- Understand the impact of new products, new pricing, or new service initiatives on customers' satisfaction, loyalty, and profitability
- Learn the products that your customers buy, how frequently they buy them, which other products they buy at the same time, and the channels they use to do their shopping, buying, and servicing.

EVALUATION FRAMEWORK FOR CUSTOMER RELATIONSHIP ANALYTIC APPLICATIONS

We've been evaluating and comparing customer-relationship analytic applications for several years. There are many products of many types being offered by a broad range of suppliers. Our experience examining these products and working with you to select, implement, and apply them has helped us refine our framework for their evaluation.

The framework has the following six sets of criteria:

- **Problem Domain.** Problem domain identifies and classifies the type and technology of the analytics in the customer analytic application product. It also identifies the type of business problem that the analytic application solves, the user roles for which the analytic application is most appropriate, and the areas of your business in which the analytic application is best applied. At the highest level, you want to understand what you're buying and what it's going to do for your business.

The types of analytics packaged in customer relationship analytic applications range from simple reporting to complex and sophisticated algorithmic analytics, such as data mining workbenches. Ideally, you should be able to apply all the types of analytics to help you understand your customer relationships. And all of the user roles in your organization that are interested in understanding customer relationships should be able to use all the types. However, we've found that specific product implementations of the types don't necessarily meet the ideal. Many don't help you understand your customer relationships, focusing instead, for example, on your internal processes. Also, many are unusable by the people in specific roles that need them (e.g., your marketing director) because they're too complex or too inflexible to adapt to your business.

- **Input.** Input includes the data that you have to provide to the analytic application and the parameters that you specify to control its execution. You don't want to have to jump through

hoops to feed data to an analytic application nor hire a PhD in statistics to run it.

- **Output.** Output includes the data produced by the execution of the analytic application and the reports generated from those data to show you the effect of the execution on your business. You want to be able to understand and apply the results.
- **Architecture.** Architecture focuses on how the analytic application integrates with your other customer-centric analytic applications. The analytic application has to fit in smoothly.
- **Viability.** Viability considers the product and the company that supplies it. You want a solid product designed for your business that is supplied and supported by a going concern.
- **Usability, Flexibility, and Customer-Centricity.** Usability, flexibility, and customer-centricity sum up our evaluations. You've made it very clear that analytic applications need these qualities to deliver the benefits that you need. You don't want to have to attend graduate school or hire consultants to use the product. You want the product to be adaptable to your business. And you need the product to promote your efforts to become a customer-centric organization. Note that we consider this criterion within our evaluation of all the other criteria and we summarize our considerations at the report's conclusion.

Before we evaluate the KXEN Analytic Framework, let us introduce you to its product modules.

KXEN ANALYTIC FRAMEWORK

The KXEN Analytic Framework has the following seven product modules:

- KXEN Event Log (KEL)
- KXEN Sequence Coder (KSC)
- KXEN Consistent Coder (K2C)
- KXEN Robust Regression (K2R)
- KXEN Smart Segmenter (K2S)
- KXEN Time Series (KTS)
- KXEN Model Export (KMX)

We can organize the seven modules into the following three groups:

- Three data preparation modules
- Three modeling modules
- A single model export module

KEL, KSC, and K2C are the data preparation modules. Each prepares a different type of input data for the modeling modules. K2R, K2S, and KTS are the modeling modules. Each is best suited to a particular type of predictive application. KMX takes models built by the modeling modules and transforms them into many different languages—such as C, Java, JavaScript, PMML2, HTML, XML, VB, SAS, ANSI SQL, DB2, Teradata UDF (SQL), Oracle UDF (SQL), MySQL, SQLServer UDF (SQL), DB2 UDF (SQL). You can use the output from KMX to score your data.

You can buy and use individual modules separately, integrating them with external analytic applications or you can buy and use multiple modules to address all aspects of one or more predictive analytic applications. For example, for the segmentation task in campaign management, you might use K2C to prepare your customer and offer data, K2R or K2S to create a classification or segmentation model that predicts which offers customers are most likely to accept, and KMX to export the segmentation model to your campaign management system.

PROBLEM DOMAIN

Data Mining

The KXEN Analytic Framework belongs in the predictive modeling or data mining problem domain for customer relationship analytic applications. Within this domain, analysts use these products and technologies to try to predict future customer behavior based on historical customer behavior and a range of other factors, including customer profile and identity characteristics.

Data mining can be applied across your business. For example, you might use data mining in marketing campaigns to predict which offers customers are most likely to accept. You can use it in sales to identify those customers who are at risk to desert or, in the lingo of the telecommunications industry, churn.

Or you can use it in service, to identify products that will cause service issues or will likely be returned.

Multistep Data Mining Process

Data mining involves the following multistep, or multiphase, process:

- **Data Preparation.** In the data preparation phase, analysts identify, collect, and organize the data that will be the input to the data mining algorithm. This phase is similar to the ETL processing of data warehousing.
- **Modeling.** A data mining model comprises a set of input data, a set of parameters for a data mining algorithm, and the data mining algorithm itself. Analysts perform modeling iteratively, adjusting the input data and the parameters for the data mining algorithm to create a model that will produce accurate, reliable, and repeatable results. Testing and training a model uses large volumes of data.
- **Scoring.** The scoring phase is the application of the model to specific data. For example, for scoring a churn model, analysts execute the model against customer data, updating (scoring) a customer attribute, likely called churn, for each customer with a value between zero and one. The higher the value, the more likely a customer is to churn.

The key strength of the KXEN Analytic Framework is that you don't need to be a data mining craftsman to deliver the benefits of data mining to your business.

Skills and Experience Required

Data mining analytic applications can generate terrific business value, but they're not widely used because, in the past, considerable skill and experience has been required to use these applications effectively. You can see this point by our descriptions of the data preparation and modeling phases of the data mining process. Data can be complex, and, historically, data mining technology suppliers haven't done much to abstract this complexity in order to make their technologies easier to use for more types of users. We've always called them data mining

workbenches—you had to be a craftsman to use them.

...But Not for the KXEN Analytic Framework

The most distinguishing characteristic and the key strength of the KXEN Analytic Framework is that you don't need to be a data mining craftsman to deliver the benefits of data mining to your business. KXEN has successfully abstracted the complexity of the data mining process. Business users will find KAF easy to learn and easy to use. In fact, we've seen implementations of the KXEN Analytic Framework in which the data mining process is performed *transparently* to users.

On the other hand, the KXEN Analytic Framework is not so "dumbed down" that analysts can't use it. The products have advanced knobs and dials that let skilled and experienced data miners fine tune input and models.

Three Roles Can Use KXEN Analytic Framework

People in the following three roles can use the KXEN Analytic Framework:

- Data mining professionals
- Business users
- IT staff

We've discussed the roles for data mining professionals and business users above. IT staff are needed for the data preparation phase to help data miners identify and collect modeling input. Many organizations have this input data within their enterprise data warehouses. Other organizations have it spread across operational systems and data marts. Because modeling will use large volumes of data, it's always best to get IT involved.

INPUT

Analytic applications typically have two types of input: data and parameters. The data will be processed by the analytic's algorithm. The parameters will control how that processing is performed.

KXEN Data Preparation Input			
Module	Input Data Types	Input Data Sources	Parameters
KXEN Event Log	Customer attribute data Transactions	Relational database tables Files	Data identification Start time End time Time period Aggregation type
KXEN Sequence Coder	Customer attribute data Behavior logs	Relational database tables Files	Data identification Start time End time Event type
KXEN Consistent Coder	Continuous, nominal, and ordinal attributes within data records	Relational database tables Files	Attribute type

Table A. This table lists the input data and input parameters for each of the three data preparation modules in the KXEN Analytic Framework.

The three data preparation modules—KXEN Event Log (KEL), KXEN Sequence Coder (KSC), and KXEN Consistent Coder (K2C)—handle the input from external sources, preparing and encoding data for analysis in the modeling phase of the data mining process. Each of the three data modules prepares a different data structure for input to the analytic modules. KEL prepares transaction data, such as those in order histories, and combines them with customer data. KSC prepares behavior data, such as those found in click streams, and combines them with customer data. K2C transforms data variables, such as age or income attributes, within any data type and storage into a form that is suitable for modeling. A byproduct of this encoding processing is the establishment of groupings for each data attribute and appropriate handling of both missing values (only if necessary) and outliers.

Easy to Use and Fast

The three data preparation modules really simplify the data preparation phase of the data mining process. Point them at the data and they do the work. What is typically an iterative process that requires detailed knowledge of input data structure and values in conventional data mining workbenches becomes a one- or two-pass process (depending on the

type of input) that any user can perform with the KXEN data preparation modules. These modules are fast, too. While we saw them handle demo data, they went through tens of thousands of records in a few seconds.

The information in Table A summarizes the input capabilities for the three data preparation modules in the KXEN Analytic Framework. Our analysis of the modules follows.

KXEN Event Log Input

You use KEL to combine data that identifies and describes customers (such as name, age, or marital status) with data about customers' transactions (such as orders, returns, and complaints). KEL combines transactional input data with descriptive input data such as customer attributes to create a set of records that is usable by any of the framework's analytic modules—K2R, K2S, and KTS.

You'd use KEL to prepare input for analyses such as predicting offers mostly likely to be accepted; the best cross-sells, up-sells, and substitutions; or future customer value.

KEL's input data can be files or relational database tables. KEL's parameters are identifiers for the two types of input data, a start time, an end time, time periods (across which to do the aggregations),

and the way in which you'd like to aggregate transaction data. Supported aggregations are Average, Min, Max, Count, and Sum. Note that KXEN plans to add other operators such as Increase (Decrease), Percent Increase (Decrease), and accumulation over different periods of time in a future version of KAF. In response, KEL combines, aggregates, and compresses the data to make them ready for any of the analytic modules in the framework.

KXEN Sequence Coder Input

You use KSC to combine data that identifies and describes customers (such as name, age, or marital status) with data about customers' behavior (such as Web click streams and call center logs). KSC combines behavior input data with descriptive input data to create a set of records that is usable by the framework's analytic modules—K2R, K2S, and KTS. (Note that we differentiate transaction data, which captures the results of behavior, from behavior data, which describes and/or represents actions such as navigating Web pages or having a conversation with a call center representative, and which may end in transactions such as an order or a return authorization. A transaction is represented in a single data record. Behavior is represented in a sequence of data records.)

The aggregation of behavior data is by event. KSC creates one input record per customer from all the events that the customer generates within the specified time interval, aggregating event transitions and generating sequencing information. That is, KSC will associate all the behavior leading up to an event with the event itself—for example, clicking the “buy” button on a Web page. KSC's input data can be files or relational database tables. This module assumes that you've already organized the behavior data using another data preparation tool. KSC doesn't build in support for these organizing capabilities, assuming, instead, that you've already got tools for this job. That's not a bad assumption, but, if it's incorrect, then you face the expense and effort of purchasing and implementing a toolset. As a result, KAF implementation can be more complicated and more expensive. KSC's parameters are identifiers for the two types of input data, a start time, an end time, and the type of event that is represented in the input data. In response, KSC combines, aggregates,

and pivots the data to make them ready for the analytic modules.

KSC makes it possible to build models that predict a customer's next activity based on a customer's previous activity. For example, you'd use KSC in order to predict which product a customer is most likely to purchase, whether the session will end after the next click, or if a chat session will need to be initiated with the customer based on his propensity to buy.

KXEN Consistent Coder

Data mining algorithms don't execute efficiently or effectively when input records contain fields (or columns when the input comes from database tables) that can have wide ranges of continuous values with outliers, non-linear relationships with the target, discrete values (with high cardinality like zip codes and product category or low cardinality like age and marital status), or missing values. K2C automates the preparation of this ordinal data using a common technique called binning (a segmentation technique that groups ranges of values into discrete buckets or bins). The module also handles outliers and plugs in appropriate values for missing values.

K2C's input data are files and relational database tables. Its parameters identify the input data sources and the input fields on which it will operate, specify outlying values, and specify replacements for missing values. K2C does binning automatically. You can see by these parameters that the data preparation tasks performed automatically by K2C can require significant manual effort and data preparation skills if you're using conventional data mining tools.

OUTPUT

Analytic applications typically have two types of output: data and reports. The data are the results produced by the processing of the analytic's algorithm on the input data. The reports present these results to analytic application users.

The output data produced by the three analytic modules in the KXEN Analytic framework are the scores, probability values, estimation, or segment indices that your models generate. For example, in analyzing your customer's likelihood to desert, you're (regression) model would generate a prob-

ability value between zero and one for each of your customers—the higher the value, the more likely a desertion. You might store this value in a customer attribute. The reports of the modeling modules show a variety of information about the model.

Three Modeling Modules

The three modeling modules—KXEN Robust Regression (K2R), KXEN Smart Segmenter (K2S), and KXEN Time Series (KTS)—produce the output for the data mining process. Execution of the models that you build with each of them scores your data, updating fields that you specify with predictions. Model execution also produces reports that present important information about model findings such as descriptive statistics, segment descriptions, and significance of attributes.

Which Modeling Approach Solves Your Problem?

Each modeling module is well suited to a particular class of modeling problem, and each expects input appropriate to the problems that it best solves. Here's where you need some data mining skill and experience. KXEN doesn't help you in deciding which modeling module to use. You have to know what problems are solved by each of its three modeling modules. That's a limitation.

With a little experience, you can figure out which modeling approach to use based on the type of results that you'd like produce. Some examples of determining the appropriate modeling approach include:

- If you are looking for cores and probabilities, use classification.
- If you are looking for estimating values, use regression.
- If you are looking for segments in your portfolio of customers or products, use segmentation.
- If you are looking to forecast future values like your sales, use time series.

- If you are looking to generate rules, use Associations. (Association Rules is a new modeling module planned for the next version of KAF.)

Once you've prepared your input, KXEN's wizard-based toolset brings you to a dialog from which you select your model building approach. The dialog asks you to select from these options:

- Classification/regression
- Clustering segmentation
- Event log aggregation
- Clustering on log aggregation
- Sequence coder
- Clustering on sequences
- Time series analysis

The event log aggregation and sequence coder selections are data preparation options. Their presence in the list only adds to the confusion. And of the three clustering options, only one, clustering segmentation, is a modeling approach. More confusion. But after eliminating them, which one of the remaining should you to choose? Clearly, you have to know something about regression, segmentation, and time series to make your choice. Let's first take a closer look at the modeling modules and see if the choice becomes more obvious.

KXEN Robust Regression

For customer relationship analytics, you can use regression to predict churn, predict how much product volume each of your channel partners will sell next year, or identify the offer that a customer will most likely accept. Regression is a widely used statistical technique that determines a value for a variable or an attribute based on known values of that variable and/or relationships of that variable with other variables. Curve fitting is the classical mathematical problem solved by regression. Knowing a set of data points, regression can find the equation for the curve that "goes through" those data points. (Linear regression fits the points with a straight line. Polynomial regression fits them with a curve.)

K2R uses a proprietary regression algorithm to build models that predict values for discrete attributes, such as offers, or for continuous variables, such as revenue or probability to churn. K2R uses K2C output to perform its proprietary implementation of

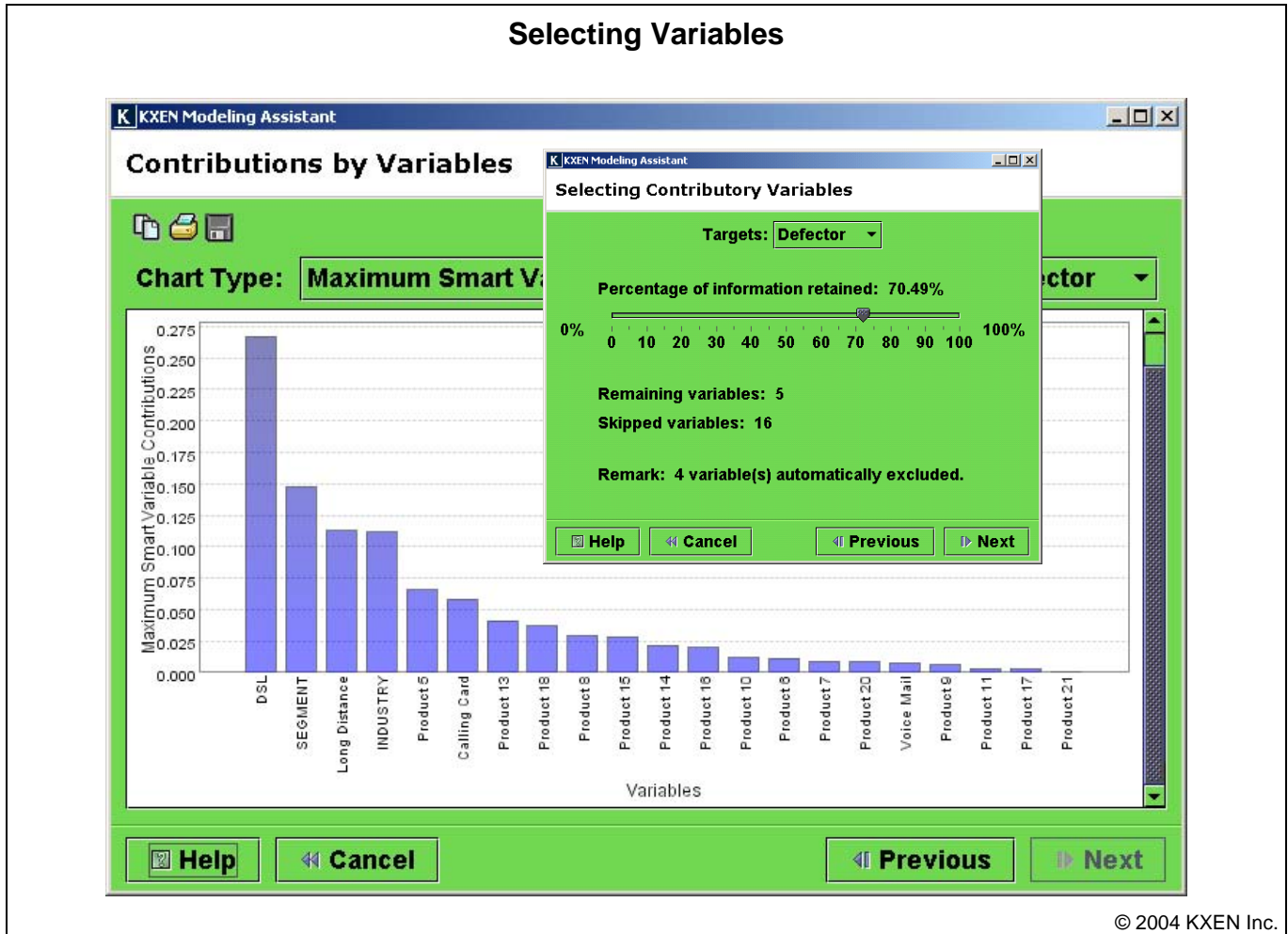


Illustration 1. This illustration shows how K2R automatically selects the best input attributes for building a regression model.

regression using an approach called Structured Risk Minimization. Internally, it maps a set of descriptive and numerical attributes (model inputs) to the target attributes (model output). The descriptive attributes come from the input that you've prepared with the data preparation modules. You identify the specific inputs to be used by K2R in the step of the wizard after you've selected "classification/regression" as your model building approach. The wizard dialog presents you with a list of attributes in input data that you include or exclude. Illustration 1 shows that dialog. K2R also requires that you specify the order of the polynomial that it will fit to your input. Most of the time, you should use a polynomial of degree one. Uh oh. That's not an ease-of-use feature. But

you can hide it from business users by setting a default for its value.

Which variable should you select? You probably don't know. You might decide to try X, Y, or Z first...That's the nub of the modeling process and one of the main reasons that the modeling process is iterative. The approach is to make an initial data selection based on your knowledge of your business, build a model with the data, run the model, examine the results, refine your input selection, and iterate. How K2R really helps here is with its speed and its reports. KXEN claims that it takes less than 15 seconds to build a model on 50,000 records containing 20 variables. You can iterate on your model-building process very quickly. However, converging on a

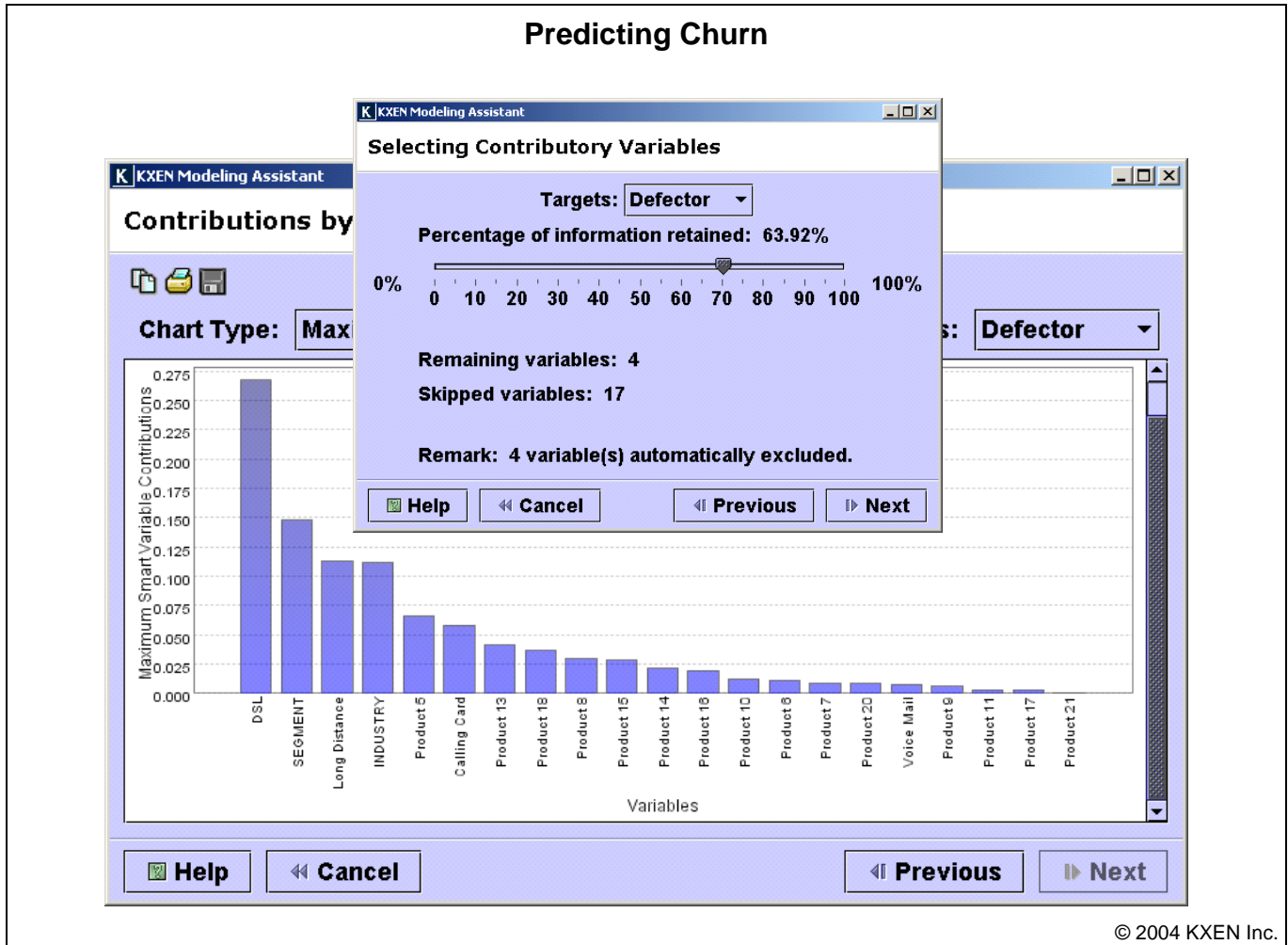


Illustration 2. This illustration shows a sample of the report produced by K2R for its modeling of churn.

high quality, reliable, and robust model is going to require selecting the right data.

INPUT. K2R can apply its regression algorithm to data prepared by KEL or KSC.

ALGORITHM. K2R uses an algorithm patented by KXEN, which is a derivation of a principle described by Vladimir Vapnik, a member of KXEN's scientific committee, as "Structured Risk Minimization." K2R automatically finds the model that represents the best compromise between quality and robustness. These models are expressed as polynomial expressions on your input variable. The only element specified by the user is the polynomial degree. To improve modeling speed, K2R can build multi-target models, models that predict multiple variables.

REPORTS. K2R reports show how each of the input attributes you select will contribute to the target attribute that you're modeling in a bar chart. The highest bar represents the attribute that's the best predictor. In addition, K2R also calculates model quality (Ki) and model reliability (Kr). These factors are indicators that a model has the ability to produce similar results on new data. Illustration 2 shows a sample report from a K2R model building to determine churn.

Once the final model has been built using the most predictive input attributes found by KXEN, you can score your data using that model and use those scores for input to your customer initiatives. For example, you might run a customer retention campaign for customers that are highly likely to defect. Or you might begin an initiative to find new

customers to replace those who will not generate much revenue for you.

KXEN Smart Segmenter

Segmentation is a process of discovering natural and meaningful groupings within a set of data. For example, within customer relationship analytics, you use segmentation to group your customers in order to determine which ones to target for a marketing campaign. Segmentation is easier to understand than regression. It's more intuitive. Also, you've probably been doing it for a while by analyzing reports or through OLAP navigation. These approaches are manual and always based on a few attributes. Use of data mining for segmentation will improve the accuracy and effectiveness of segmentation, because data mining segmentation will use many more input attributes and make this process more automated.

INPUT. K2S can discover segments from data prepared by KEL or KSC.

ALGORITHM. Segmentation processing can either be supervised (because a target attribute is specified) or unsupervised (because no target attribute is specified), depending on the availability of a business question, which specifies a target attribute. KXEN recommends the supervised approach. K2S uses a clustering algorithm that is distance based in order to create initial clusters. It then applies a mathematical algorithm to find minimal descriptions for each cluster. These descriptions are rule-based expressions that cover the entire population with minimum overlap. K2S, like all the modeling modules, is very fast. You can iterate your modeling rapidly with large numbers of input variables and large data volumes. A financial institution uses KXEN to segment its customer base using 2,500 input attributes they have collected on their customer base.

REPORTS. The reports generated by K2S show how well the input variables that you specify predict segmentation. And, as with all the modeling modules, K2S also calculates model quality (Ki) and model reliability (Kr). Illustration 3 shows a sample report from K2S model building.

Once you're satisfied with your segmentation model, you can execute the model to score your customer data. The scoring process for K2S assigns a segment identifier to each input record.

KXEN Time Series

You can use KTS to forecast customer growth, to identify seasonal changes in orders, or to identify the periodicity in customer service activity. KXEN Time Series performs forecasting. It takes historical time series data as input and predicts values for future time periods. KTS also identifies and predicts seasonality and periodicity in time series data. Nothing mysterious here. Most of you already do some sort of forecasting.

INPUT. KTS can use data prepared by KEL or KSC, however this is rarely needed. Most commonly, KTS uses time series data that you prepare.

ALGORITHM. KTS is based on a multi-stage K2R approach that finds the best combination of regression models to explain the trends, the seasonalities, and the cycles in input data. After finding these components to describe the data, KTS works on the residuals, the data that are not captured by previous regression models. The mathematical approach used is a proprietary implementation of ARIMA models (a seasonal adjustment program for time series data from the U.S. Census Bureau).

REPORTS. For time series, KXEN generates a decomposition report that describes the different components of the model: the trend, the seasonality, the cycles, and the residuals.

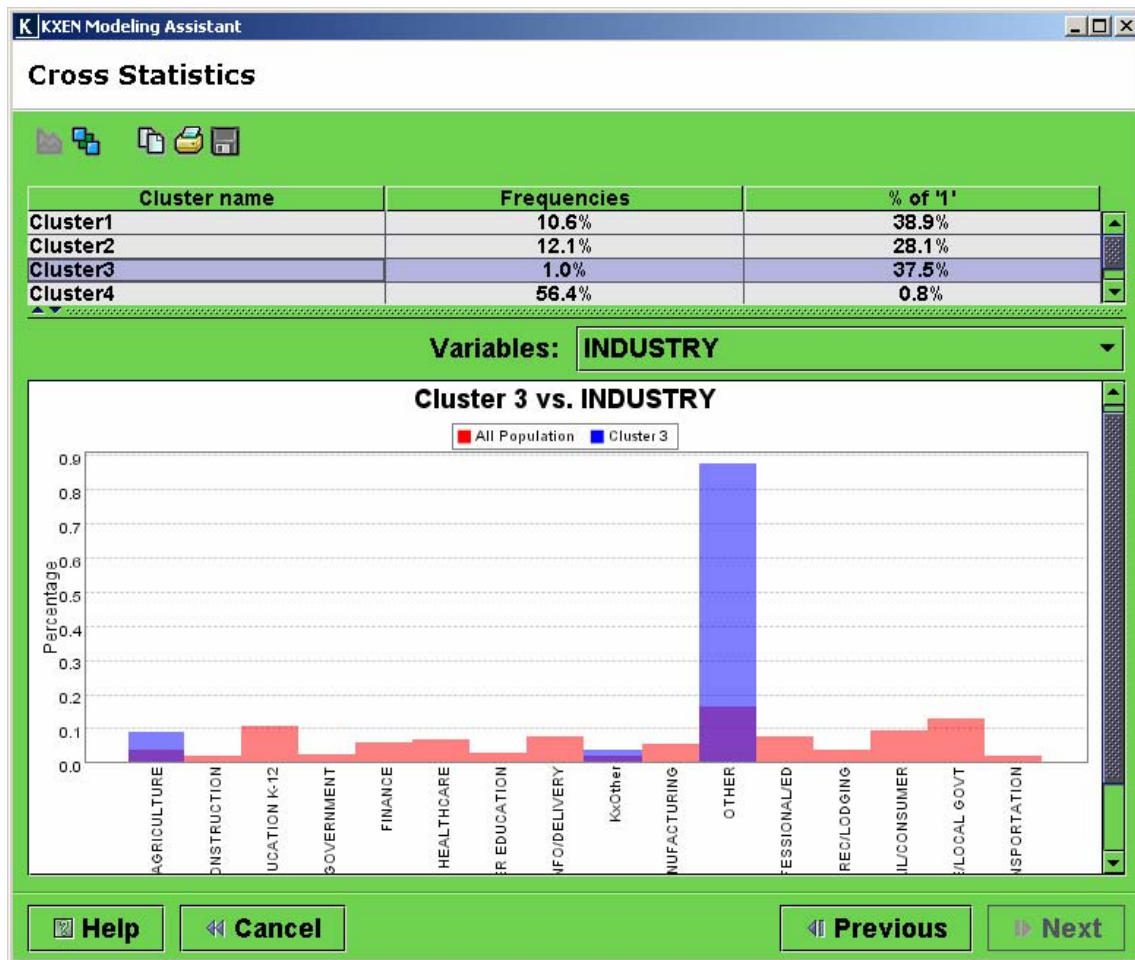
ARCHITECTURE

Architecture defines how products perform analytic application functionality. Architecture is not as important for customer-centric analytic applications, which typically have a few users who work together in a single location, as it is for operational systems or analysis/reporting systems with large user communities.

We've usually looked at six architecture evaluation criteria: environments, organization, infrastructure, structure, customization, and integration. For customer-centric analytic applications we'll consider only the following three:

- **Environments**, which are the server platforms and databases that a campaign management product supports. If a customer-centric analytic application doesn't support your existing envi-

K2S Segmentation



© 2004 KXEN Inc.

Illustration 3. This illustration shows a sample of a report produced by K2S.

- ronments, then its implementation probably won't be worth the cost and effort required to acquire the environments that it needs.
- **Organization**, which characterizes a product's architecture and identifies a product's major components and the interfaces between them. (Interfaces are always between two components.) Organization comes into play for the reporting output of customer-centric analytic applications. While few users are needed to create input and run the application, you'd like to involve a wider community in reviewing results. So some sort of Web-based reporting is needed.

- **Integration**, which complements and completes processing with the functionality of external applications. You use customer-centric analytic applications to understand and refine your operational systems. Refinement means that you'll have to use the results of analytic processing to change operational processing. That means integration.

KXEN Analytic Framework Architecture

The environments, organization, and integration of the KXEN Analytic Framework are described in Table B.

KXEN Analytic Framework Architecture	
Criterion	Description
Environments	<p>Server platforms</p> <ul style="list-style-type: none"> • IBM AIX • Hewlett Packard HP-UX • Linux • Microsoft Windows • Sun Solaris <p>Databases</p> <ul style="list-style-type: none"> • Files including text files, SAS file formats, SPSS files, Excel files, and proprietary storage formats • IBM DB2 • Microsoft SQL Server • Oracle Server • Teradata RDBMS
Organization	<p>KXEN Analytic Framework has a flexible organization. You can deploy it as a stand-alone system, a two- or three-tiered client/server application, or a Web application. KAF provides clients and application server, but it does not provide its own data-base tier. Clients are the native clients of the supported platforms and a Java client.</p>
Integration	<p>KXEN Analytic Framework integrates with external applications through published APIs. KXEN provides API bindings in DCOM for Windows platforms and in CORBA and C++ for all server platforms. It also provides a Java API on top of these bindings.</p> <p>Integration can also be accomplished by accessing the KXEN Analytic Framework through commands. APIs for its command interpreter are also published, but only C++ bindings are supported.</p> <p>In addition, KXEN Model Export (KMX) lets you export KXEN Analytic Framework Models to be executed within other analytic application environments. These include C/C++, Java, JavaScript, PMML2, HTML, XML, VB, SAS, ANSI SQL, DB2, Teradata UDF (SQL), Oracle UDF (SQL), MySQL, SQLServer UDF (SQL), and DB2 UDF (SQL).</p>

Table B. This table describes the key aspects of the KXEN Analytic Framework's architecture.

KAF's architecture is quite flexible but a little complex. Integration is the strength of the architecture. Multiple integration approaches, a wide range of environments supported for model export, the three API bindings (plus Java on top of them), and the multiple forms of organization supported make implementation of the framework extremely flexible. In fact, KXEN designed the framework more to be integrated within external applications than to be used independently. So we're not surprised at the lack of a data tier or at the somewhat clunky native UI. Note, however, that KXEN is planning a new UI for a future version. Also, integration facilities are rather low level. That's not a disadvantage for ISVs that will integrate KXEN into their offerings, but C++, CORBA, and DCOM integrations are low level and not the most modern. Note that here, too, KXEN plans improvement. The firm's R&D organi-

zation is working on the reference implementation of the JDM (Java Data Mining) API, which will enable standards-based and modern integration.

PRODUCT VIABILITY

Product marketing criteria consider the business aspects of campaign management applications and their suppliers. These criteria are much easier to evaluate than functionality criteria, but they can be deal breakers. The seven product marketing criteria we consider in evaluating an offering are:

- Product background
- Target market(s)
- Installed base
- Price
- Product plans

- Competition
- Company viability

Product Background

KXEN Analytic Framework is a five-year-old product that has been delivered, maintained, and enhanced through multiple versions. Across those versions, KXEN has continually and significantly improved it. We have no significant product viability concerns about the KXEN Analytic Framework.

KXEN Analytic Framework 3.0 is the third and current version of KXEN's only product offering. 3.0 was introduced in June 2003. The first product version, 1.0, was introduced in 1999 and included the Event Log (KEL) and Smart Segmenter (K2S). 2.0 was introduced in June 2001. This version added Robust Regression (K2R), Consistent Coder (K2C), and C Code Generator (KCG) (a module that generates data mining models in C code). KXEN added Sequence Coder (KSC), Model Export (KMX), and Time Series (KTS) in 3.0.

Target Market

KXEN has a multitiered target market for its analytic framework. At the top tier, the firm targets the following three types of organizations:

- OEM partners such as Business Objects and DoubleClick
- System integrator partners such as Accenture and IBM
- End users in finance, telecommunications, and retail

Within its end user target market, KXEN targets three industry segments: finance, telecommunications, and retail. At the lowest level, the firm targets churn and decision-making applications within the telecommunications segment of end-user organizations. These are the markets in which data mining has been most successfully used.

Installed Base

KXEN claims that more than 200 customers have installed modules from its Analytic Framework. Across its target markets, here are a few examples of

reference customers: Business Objects, DoubleClick, and Clarity Blue are OEM customers. We evaluated the Customer Analytics module of BusinessObjects Customer Intelligence against the framework in this report in February 2004. All of the predictive analytics in this customer relationship analytic application are built on the KXEN Analytic Framework and, as we mentioned earlier in this report, the details of the data mining involved in those analytics are completely transparent to users. DoubleClick embeds Smart Segmenter functionality to help do segmentation in its campaign management offering.

Accenture, Experian, and Teradata are customers in the system integrator partner target market. Teradata and KXEN have integrated the KXEN Analytic Framework with Teradata Warehouse. What's interesting about this partnership is that Teradata had developed its own data mining product, Teradata Warehouse Miner, but it still partnered with KXEN to be able to offer a faster and more usable data mining capability.

KXEN has some well-known reference customers among end users, especially in banking and telecommunications. These customers include Barclays, France Telecom, and JP Morgan Chase.

Price

You can buy the individual modules of the KXEN Analytic Framework for an initial license fee plus an annual maintenance fee, or you can rent them for a monthly rental fee. The license has a per CPU per product module pricing model with 18 percent annual maintenance. You can also buy product module bundles. For example, for \$125,000, you can license the Sequence Coder (KSC), Robust Regression (K2R), and Smart Segmenter (K2S) on four CPUs.

KXEN Analytic Framework is attractively priced. License fees are reasonable and, given its usability and broad range of applications, you should be able to get a good return on your license investment pretty quickly.

Product Plans

KXEN has plans to improve all aspects of KAF. A new modeling component, KXEN Association Rules (KAR), will support an associative rules algorithm. In addition, data preparation and transforma-

tion capabilities will be added. KXEN is also developing a new user interface. In the area of architecture, KXEN is developing the reference implementation of JDM-API (JSR-73). As a result, KXEN will deliver the first Web Services associated with this standard. And in the area of integration, KXEN is working on integration with the major ERP suppliers and OLAP vendors.

These plans really strengthen the offering. Unfortunately, they don't address the limitations that we've identified, particularly in choosing the right modeling approach.

Competition

You can buy data mining products from many types of data mining technology suppliers. All of them compete with KXEN.

DATA MINING WORKBENCHES FROM DATABASE SUPPLIERS. Most of the major database suppliers offer data mining capabilities as features. For data warehousing applications, these suppliers typically bundle the data mining capabilities within a package that also includes the database and data warehouse building and management tools. We'd classify these data mining database features as workbenches—low-level toolsets that require data mining and data preparation skills and experience to prepare input data, build and test data mining models, and generate scores with selected models. Their major advantages are their packaging and their integration with the database that provides input. Some of these offerings, IBM's Intelligent Miner for DB2 in particular, also have a very wide range of data mining algorithms. These competitors are also all large and well-established companies that have company viability advantages over KXEN. KXEN has significant usability, abstraction, and performance advantages over all these offerings.

ESTABLISHED ANALYTICS SUPPLIERS. KXEN also competes with established analytics firms like SAS and SPSS. These competitors offer low-level data mining workbenches as well as more usable analytic applications that embed data mining technology and other analytic technology such as optimization. But they are also beginning to offer more usable analytic applications that abstract the complexities of data mining, competing more closely

with KXEN. These competitors are also larger, more established companies than KXEN. However, as against database suppliers, KXEN still has usability, abstraction, and performance advantages over most offerings of these competitors.

CAMPAIGN MANAGEMENT SUPPLIERS. We've evaluated a few campaign management products that integrate and/or add on data mining capabilities to help with segmentation and offer targeting. Unica Affinium and E.piphany Marketing come to mind. Products of this type also compete with KXEN Analytic Framework, and these products have the advantage of tight integration. Their usability can be pretty good, too. E.piphany is especially strong in usability. KXEN fights back pretty well both indirectly and directly. DoubleClick is an indirect example. As we mentioned above in "Installed Base," DoubleClick integrates Analytic Framework within its campaign management offering for segmentation. Directly, KXEN's end-user customers use the Analytic Framework within marketing. Its usability and performance are an edge over most competitors.

SMALL DATA MINING SUPPLIERS. KXEN competes most closely against small data mining suppliers—companies similar to itself. Coincidentally, these competitors are all European companies. Norkom, a company based in Ireland, for example, takes a more applications-oriented approach toward usability. Its Alchemist product has anti-money laundering, fraud detection, risk management, business performance management, and marketing automation applications. Quadstone, a U.K. company, has a customer focus for its data mining products: Decisionhouse, Transactionhouse, and Actionhouse. Quadstone's products simplify the application of data mining, but these products are designed for analysts. Quadstone also does a good job with data preparation. KXEN Analytic Framework has usability advantages over many of these competing products.

Company viability

KXEN is a privately owned company incorporated in Delaware. It was founded in France in 1998. The firm is currently headquartered in San Francisco, California, with research and development facilities in Paris, France. Its vision is "to make data

mining part of your every day business decisions.” Its mission is “to embed advanced analytics into existing enterprise applications and business processes.” In achieving this mission, KXEN lives up to its name, an acronym for Knowledge eXtraction ENgine.

KXEN is as much about the science of analytics as it is about the analytics business. The company has a chief science officer as well as a chief technology officer. KXEN also has a scientific committee, whose members offer guidance on product technology. Its members are some of the world’s best minds in statistics, including Vladimir Vapnik, Gregory Piatetsky-Shapiro, Lee Giles, Gilbert Saporta, Yann Le Cun, and Bernhard Schölkopf. For example, Vapnik, who was appointed in April 2003, is an expert in machine learning and theoretical and applied statistics. Professor Vapnik has developed a general theory for minimizing the expected risk of losses using empirical data and a new type of learning machine called the Support Vector machine. His work has been used to solve pattern recognition and regression estimation problems. This work closely aligns with KXEN’s products and the business problems that they solve. KXEN’s scientific bent ensures that its products incorporate the best and latest developments in statistics. This approach has significant business advantages. The advanced statistical techniques that its products incorporate enable you to handle more data and produce usable results faster for more types of problems than most data mining approaches.

KXEN’s approach to analytics has worked very well. This is a small and young company that has had significant success in a short period of time, developing and offering products built on technology that continues to be hard to sell. And they’ve done it primarily in tough economic times. KXEN’s formula for combining the best science with the best usability has truly made data mining a tool for “every day business decisions.”

EVALUATION AND CONCLUSION

Customer data mining has long been a valuable analytic tool. Companies that have been able to afford the investments in staff and time necessary to develop the expertise required to use customer data

mining effectively have been getting very good paybacks. For example, retail banks are very effective in their segmentation, cross-selling, and up-selling through the use of data mining. However, with customer bases in the millions, these organizations had to find a way to do the analyses that most organizations can do with historical reports and OLAP.

Speaking of OLAP, that brings us to you. We urge you to consider data mining. With it you can perform consistent, automated, and predictive analysis of your customer relationships. Data mining is so much more powerful than reporting techniques. That power can make your analyses more effective and more efficient. More effective and more efficient analyses drive business improvements like more profitable customers.

KXEN Analytic Framework breaks down the historical objections that you may have had toward data mining. It reduces the complexity of data mining, making it easy to learn and easy to use. It also makes data mining fast, speeding both the process and the processing.

Bottom Line

If you’ve been resisting data mining because you thought it was too complex, if OLAP isn’t giving you consistent answers on a timely basis, or if you’ve been struggling to continue to improve your customer relationships after an initial big hit, then KXEN Analytic Framework might be the analytic toolset for you. It truly does make data mining technologies more accessible and more usable. Data mining can help you understand and predict key factors that affect your customer relationships. It’s time to expand your analytic toolset beyond OLAP. The KXEN Analytic Framework is a next step in analytic applications that you should consider taking.

Contact Info:

KXEN Incorporated
201 Mission Street, Suite 1950
San Francisco, CA 94105-1831
Joerg Rathenberg
Vice President Marketing & Communications
Phone: 415.904.4168
Email: Joerg.Rathenberg@kxen.com
Internet: <http://www.kxen.com>