

Advanced Data Mining & Forecasting

KEY BENEFITS

- **Quickly add data mining and forecasting capabilities using embeddable libraries**
- **Select from a broad range of classic and advanced techniques**
- **Solve unique data challenges with customized techniques**
- **Integrate numerically accurate and fully documented libraries**
- **Visualize virtually any dataset**
- **Reduce risk by embedding commercially tested and supported technology**
- **Leverage decades of proven expertise across a wide range of industries**

Challenges and Opportunities

Organizations today capture greater quantities of data about customers than ever before. Automated applications make it possible to capture multiple layers of customer characteristics, while inexpensive storage makes it possible to save this data for long periods of time. For organizations, a major challenge – as well as a major opportunity – is how to effectively use all of this data to make beneficial business decisions.

Solid data mining and forecasting abilities benefit organizations across a variety of fields and in many application areas. Whether forecasting future expenses, planning manufacturing production times, predicting sales trends or deriving an optimal trading strategy for securities, making data-driven decisions can improve revenue, reduce costs or mitigate risks.

Different data mining techniques perform better than others depending on the information that organizations hope to capture from the data as well as the types and quantities of the data itself. To effectively address data mining and forecasting challenges, analysts need flexible, scalable and reliable analysis tools.

Unique Analysis and Visualization Tools

Visual Numerics offers data mining and forecasting solutions through a unique toolset of products and services.

- The IMSL[®] Numerical Libraries, available in C, C# for .NET, Java[™], Fortran and as a Python interface to the C library, offer a wide range of numerical analysis techniques for data mining, forecasting and advanced predictive analysis. The IMSL Libraries contain classic approaches to forecasting and predictive analysis such as regression analysis, correlation analysis, cluster analysis, analysis of variance and interpolation. The IMSL Libraries also offer a number of more advanced techniques such as covariance analysis, discriminant analysis, neural network and naïve bayes classification, genetic algorithm optimization, and a comprehensive range of probability distributions and random number generators that can be used for Monte Carlo simulations.
- PV-WAVE[®] works standalone or integrated with the IMSL Libraries to provide sophisticated visualization of forecasting models. The PV-WAVE Family of products offers a highly flexible and scalable development environment for 2D, 3D to nD visualization, a Java-based solution for networked applications, and an advanced time series analysis solution. The combination of these tools and their integration with the IMSL Libraries provides unsurpassed analytic and visualization power to users for a wide range of forecasting techniques and other scientific and business applications.

Whether using the IMSL Libraries, PV-WAVE or other software solutions, Visual Numerics' team of Consulting Services experts have decades of analytics experience to help customers research, define, and implement data mining and forecasting applications. Visual Numerics' approach is to build solutions to customers' specific needs and data, versus watering down data to fit into a standard format, which can compromise a customer's competitive advantage.

SAMPLE TECHNIQUES AVAILABLE FROM VISUAL NUMERICS

- **Regression**
 - Linear and nonlinear
 - General least squares
 - Multiple variables
- **Cluster Analysis**
 - Hierarchical
 - KMeans
- **Neural Networks**
- **Auto_ARIMA**
- **ARMA, GARCH**
- **Genetic Algorithm**
- **Naïve Bayes**
- **Logistic Regression**
- **Principal Components Analysis**
- **Factor Analysis**
- **Variances & Covariances**
- **Discriminant Analysis**
- **Analysis of Variance**
- **Visualization**

A few algorithms are only available in specific IMSL Libraries. Contact Visual Numerics for details.

Classic Techniques

In many applications it is useful to identify whether new data fits expected patterns based on historical data. For example, a manufacturer might want to identify when the number of defects is higher than expected. Basic cumulative distribution functions like chi-squared, student's t and F distributions are useful in these applications.

Other classic data mining techniques like regression analysis and time series techniques allow users to easily model any underlying data structure. These techniques are typically mathematically straightforward but perform quickly and provide diagnostics and insight for forecasting. Cluster analysis is another classic technique that enables analysts to partition data into various subsets.

• **TERADATA STATISTICAL LIBRARY POWERED BY IMSL**

To provide users with a wide range of embedded statistical functions for in-database analysis, Teradata embedded multiple IMSL Library probability distribution functions, inverses and random number generators into a new product, the Teradata Statistical Library Add-in, Powered by IMSL, part of Teradata Warehouse Miner.



These statistical functions are provided as embedded user defined functions (UDFs) accessible through Teradata Warehouse Miner or standard SQL Queries. Teradata Warehouse Miner's UDF capabilities work like an index from which users can select prepackaged functions which are linked to the detailed data. From a desktop environment, business users can click, drag and drop the needed analytic function from a list of UDFs and run it against the appropriate detailed data in the warehouse without wasting time by manually coding the function or moving data between systems. Combined with other analytic functions in Teradata Warehouse Miner, the Statistical Library functions, Powered by IMSL can be used in a variety of applications, including classification, forecasting and others.

EXAMPLE

Sophisticated Neural Network Technology

Neural network forecasting and classification functions help users discover relationships and valuable information in vast amounts of data. Neural Networks provide a highly flexible forecasting approach that delivers accurate forecasts under a variety of conditions including complex data systems, short time series, mixed categorical and continuous data, noisy data and a large number of variables.

Visual Numerics' Neural Network technology features advanced data pre-processing to simplify data mining preparation. These automatic pre-processing algorithms are tailored to work with the Neural Network training engine to ensure optimal accuracy and performance, saving significant time over manual pre-processing. Unlike other neural network algorithms, the Neural Network training algorithm implementation in the IMSL Library provides control over parameters such as the number of training epochs, specific linkages within the network and activation functions. In addition, there is no limit to the number of layers and perceptrons that can be used in the network. This high degree of flexibility and control is especially beneficial to analysts in the fields of finance, business analytics, bioinformatics and life sciences, where data sets can be extremely large and complex.

• **TEKSOUTH: ADVANCED PREDICTIVE ANALYTICS SOLUTION FOR MORE ACCURATE BUDGET FORECASTING**

To enable private sector and government organizations to more accurately project and budget spending changes, and help companies address accountability and regulatory compliance issues, TekSouth, working with Visual Numerics Consulting Services, developed an innovative, customized neural network forecasting application for its customers based on neural network forecasting functionality in the IMSL Library.



EXAMPLE

The application extracts information from historical cost data, applying the information to forecast future costs. The analytic engine has the flexibility to prioritize historical data and exclude anomalous data to improve its predictive capabilities. One of the most powerful features of this solution is its ability to mimic the brain's own problem-solving process by applying knowledge gained from historical data to new problems, fine-tuning its forecasting accuracy even more over time.

"The flexibility and learning potential of our neural net forecasting application has significant potential to help government agencies and commercial businesses improve their predictive analysis capabilities, moving well beyond traditional tools into high end business intelligence."

VICE PRESIDENT OF STRATEGIC PLANNING AND CORPORATE DEVELOPMENT FOR TEKSOUTH, AND A RETIRED U.S. AIR FORCE MAJOR GENERAL

"With VeriSight, we combine massive amounts of healthcare data, such as medical and pharmacy claims, in a highly secure and confidential environment. JWAVE gives us the performance and powerful image display capabilities we required for valuable, real-time visualization of that data."

VICE PRESIDENT AND CHIEF TECHNOLOGY OFFICER HUMANA

Auto_ARIMA Advanced Forecasting Routine for Time Series Analysis

The IMSL Library function, Auto_ARIMA, is an advanced forecasting routine for time series analysis with an ARIMA model. "Auto" as in automatic, refers to the fact that unlike traditional functions, Auto_ARIMA requires very little, if any, user pre-processing of the time series data.

ARIMA is a powerful and flexible methodology, but it is an extremely complex technique and requires a great deal of experience to be used effectively. The accuracy of results heavily depends on the analyst's level of expertise. With Auto_ARIMA, the complexity of time series analysis with ARIMA is greatly simplified due to its accurate and automatic pre-processing of data, allowing for faster time to results.

The Auto_ARIMA function is applicable for applications such as sales forecasting, commodity pricing (e.g. oil & gas), stock market predictions, semiconductor yield analysis, and more. Utilizing the expertise of Visual Numerics' Consulting Services Team, an organization can integrate Auto_ARIMA to create an optimal and custom forecasting solution for their situation.

• MAJOR EUROPEAN TELECOMMUNICATIONS COMPANY: ADVANCED FORECASTING FOR CELLULAR NETWORK OPTIMIZATION

For a major European telecommunications company, planning and managing their cellular network is critical for their success. Too little coverage and customers experience service problems; too much coverage and resources are wasted. Good planning means analyzing the massive amounts of usage data generated by the thousands of cellular towers in the service provider's network. To track usage on their German mobile network, the service provider already utilized sophisticated data analysis techniques, but needed another level of refinement in order to better understand traffic patterns.



For a company owning thousands of cellular towers, all sending data to their Network Operations Center, traditional statistical analysis techniques were inadequate, labor intensive, expensive and cumbersome. The service provider turned to Visual Numerics Consulting Services group. To identify the outliers and locate any seasonal trends in the data, Visual Numerics leveraged the ARIMA algorithm, automating the algorithm to tackle the customer's faster processing requirement. The state-of-the-art, expert-system Auto_ARIMA algorithm allowed the service provider to produce more accurate forecasts faster than before, giving them a significant advantage in the market.

EXAMPLE

Genetic Algorithm for Advanced Search and Optimization

Genetic Algorithms, originally introduced in the 1970s by John H. Holland and his students at the University of Michigan, are increasingly popular for solving optimization, search, machine learning and pattern recognition problems. A genetic algorithm solves problems in an evolutionary fashion, first developing a set of possible solutions, then identifying the best solutions from that set and recombining them to form new solutions. The process repeats, with poor solutions being replaced by better ones, until an optimum solution is identified.

A genetic algorithm can provide valuable functionality for many data mining applications. For example, by identifying the best indicators that will determine if a credit card applicant will be a credit risk, or by identifying patterns in purchase behavior to enable companies to better target price discounts.

Naïve Bayes for Text Mining and other Large Classification Problems

Naïve Bayes is a simple algorithm that is very fast. A Naïve Bayes classifier can be trained to classify patterns involving thousands of attributes and applied to thousands of patterns. As a result, Naïve Bayes is a preferred algorithm for text mining and other large classification problems.

As most organizations today have more text and documents than humans can keep track of, text mining is becoming an increasingly popular data mining and forecasting tool. With the IMSL C Library Naïve Bayes algorithm, developers can create applications that search websites or customer relationship management system notes for entries that indicate unsatisfied customers or create applications that find materials on the web related to other materials that have proven valuable in the past (e.g., identify new financial news feeds that will provide useful indicators for a specific portfolio).

Visual Data Analysis

Whether used on its own or with the IMSL Libraries, the PV-WAVE Family of products provides a useful set of visualization tools for data mining and forecasting. With its complete set of flexible graphics functions for developing both standard and custom views of data. The PV-WAVE scripting language allows for the rapid development of prototypes and quick turnaround for business user requests, while the array based language makes it simple for users to drill-down on data.

- **HUMANA: SOPHISTICATED CLINICAL HEALTH METRICS APPLICATION**

When Humana, a large health benefits company needed advanced visual data analysis for its VeriSight™ application, they selected JWAVE, part of the PV-WAVE Family. VeriSight allows Humana analysts and clinicians to identify trends in member data, to analyze the impact of existing programs and products, to identify new service opportunities, and to supplement risk management models.



Using visual data analysis techniques developed with JWAVE, Humana's VeriSight users identify relationships among the myriad variables of the member populations served by Humana to anticipate future costs, clinical events and guidance needs. Users are able to render secure intranet Web pages on-demand, enabling faster, more interactive exploration of clinical program effectiveness and participation, determination of cost drivers and other analytics that fuel programs geared to improving member health and to driving down costs.

EXAMPLE



Visual Numerics Corporate Headquarters

2500 Wilcrest Drive, Suite 200
Houston, TX 77042

USA Contact Information

Toll Free: 800.222.4675
Houston, TX: 713.784.3131
Broomfield, CO: 720.407.4200
Email: info@vni.com
Web site: www.vni.com

Visual Numerics has Offices Worldwide

USA • UK • France • Germany
The Netherlands • Mexico • Japan
Korea • Taiwan • China

For contact information, please visit
www.vni.com/contact