M2006

Knowledge that makes a difference

NINTH ANNUAL DATA MINING CONFERENCE

October 23-24 **Las Vegas NV**

Conference Program

§sas | THE POWER TO KNOW

Dear Data Mining Colleague,

As conference co-chairs for M2006, we'd like to be the first to welcome you to Las Vegas for SAS' ninth annual data mining conference.

Given the amazing growth data mining has enjoyed over the last decade or so, we think it's appropriate that we're gathering in the nation's fastest growing city. When we started this conference back in 1998, we attracted around 100 attendees and a dozen speakers. Today we find ourselves surrounded by several hundred data mining practitioners and more than 40 invited speakers. Clearly, our growth is a testament to the field's dramatic expansion.

In the next two days you'll discover how various companies across virtually every industry use data mining to improve their organization. You'll learn the latest academic research and discover business best practices that can be taken back and implemented at your organization. However, the knowledge you'll discover during formal presentations is only part of the benefits you'll realize at M2006.

While we're certain you'll enjoy the outstanding agenda we've developed, we ask that you also take advantage of the numerous networking opportunities available to you. M2006 has attracted data mining professionals from nearly twenty countries and three hundred different companies. We hope you'll also make full use of the opportunity to learn from your peers.

Once again, we extend a warm welcome to you. Please enjoy the conference and let us know if we can assist you at any time during the event.

Jerry Oglesby
M2006 Conference Co-Chair
SAS

Tom Bohannon
M2006 Conference Co-Chair
Baylor University

Inside the Conference Program

knowledge that makes a difference

Meet the Co-Chairs

Jerry L. Oglesby holds a Ph.D. in Statistics from Texas A&M University, M.S. in Mathematics from Lamar University, and a B.S. in Mathematics from the University of Mississippi. He currently works for SAS as the Director of Higher Education Consulting within the Education Division. This department is charged with supporting SAS in the university wide community. It has, as its major goal, the introduction of SAS training materials and software in the curriculums of courses across many units within the universities. Prior to starting this group, he was Director of Analytical Consulting within the Professional Services Division. As Director of Analytical Consulting he grew the Department from its formation to approximately forty modelers and business analysts whose primary function was to provide analytical support and expertise to SAS' sales force and customers. This group was largely responsible for the support of the successful launch of SAS' award winning data mining solution, SAS® Enterprise Miner™.

From 1990 until joining SAS in July of 1996, Jerry was employed by Monsanto Chemical Company as plant statistician and Manufacturing Technologist. He was CEO and founding President of SCI Data Systems from 1977 to 1990. Following completion of his doctorate at Texas A&M in 1971, he was a professor of Statistics at the University of West Florida where he established the Institute for Statistical and Mathematical Modeling for doing analytical and computational consulting for clients on and off campus.

Tom Bohannon, Ph.D. is the Assistant Vice President and Director at the Office of Institutional Research & Testing Services at Baylor University. The primary mission of the office is to engage in continuing research in all pertinent aspects of Baylor University and to provide complete testing services for Baylor students and public. Prior to this position, Dr. Bohannon served as the Director of the Office of Institutional Research at Appalachian State University and as an Associate Professor of Mathematical Sciences at Appalachian State University.

Dr. Bohannon has served in leadership positions in several professional organizations including AIR, SAIR, NCAIR, and TAIR. He has presented papers and workshops at the following professional associations for over 20 years—AIR, SAIR, NCAIR, TAIR, ASA, SUGI and others. His areas of specialization include Statistics, Data Warehousing and Data Mining, Institutional Research and Computer Application, SAS Training and Applications, Microcomputer Training Applications.

Dr. Bohannon earned a Ph.D. in Statistics from Texas A&M University in 1976 and an M.A. in Mathematics from Wake Forest University in 1965. He also holds a B.S. in Mathematics with a Physics minor from McNeese State University.

Monday, October 23

7:30 a.m.	**Registration open** *Continental breakfast available in the Exhibit Hall*	**Salsa Registration Pavilion 7**
8:45 a.m.	**Welcome from Co-Chairs**	**Pavilion 9**
9:00 a.m.	**Keynote Address** **A View of the Data Mining Revolution and Evolution** William (Bill) Smith, Executive Director American Statistical Association and Professor Emeritus of Statistics, Texas A&M University	**Pavilion 9**
10:00 a.m.	**Break** *Refreshments served in the Exhibit Hall, sponsored by Bank of America*	**Pavilion 7**
10:30 a.m.	**Keynote Address** **From Data Mining Grand Challenges to the New Sciences Underlying the Internet** Usama Fayyad, Chief Data Officer and Vice President of Strategic Data Solutions Yahoo! Inc.	**Pavilion 9**
11:30 a.m.	**Lunch** *Exhibit Hall Open*	**Amazon A**
12:30 p.m.	**Keynote Address** **Building a Data Mining Community** Dave Duling, Software Development Manager for SAS® Enterprise Miner™ SAS Institute Inc.	**Pavilion 9**
1:30 – 2:30 p.m.	**Breakout sessions**	
	Perspectives from Healthcare **Mining Without a Hardhat: A New Approach to Safety at a National Laboratory** Judy Spomer, Knowledge Discovery and Extraction Sandia National Laboratories	**Palma ABEF**
	Business Applications **Why Data Mining is an Encapsulating Solution to Querying, BI and BA in Research and Planning in Higher Education and Beyond** Jing Luan, Vice Chancellor, Educational Services and Planning San Mateo CCCD	**Palma CDGH**
	Business Applications **Needs Based Segmentation and its Application in SKTelcom** Jongdo Kim SKTelecom, Korea	**Amazon K, L, M**
	Fraud Detection **Detecting Fraud by Building Strong Control Environments using Continuous Monitoring** William (Bill) J. Kelley COL FA Data Mining Division, Department of Defense Office of Inspector General	**Pavilion 3**
	Perspectives from Retail **Using SAS Data Mining Procedures to Estimate "True" Value of New and Existing Products** Abdu Elnagheeb, Senior Vice President Bank of America	**Pavilion 4**
	Data Mining in Marketing **Predictive Modeling and Segmentation in the Newspaper Industry** Goutam Chakraborty, Ph.D., Associate Professor of Marketing Oklahoma State University	**Tropical A-D**
	New and Emerging Technologies **Visualizing Multiple and Logistic Regression Models** George Fernandez, Professor of Applied Statistics University of Nevada, Reno	**Tropical E-H**

2:45 – 3:45 p.m.	**Breakout sessions**	
	Perspectives from Healthcare **Using Administrative Records to Assess the Performance of Healthcare Providers—Pitfalls and Challenges** HungChing Chan, Director of Clinical Analysis and Program Evaluation Medica	**Palma ABEF**
	Business Applications **Customer Acquisition and Retention using Data Mining Techniques** T. Lynn Locke, Director of Database Marketing Blue Cross Blue Shield of Florida	**Palma CDGH**
	Fraud Detection **Data Mining at the Texas State Auditor's Office** Tom Winn, Senior Systems Analyst, Texas State Auditor's Office	**Pavilion 3**
	Perspectives from Retail **Customer Profit Value in the Insurance Business** Günter Schmölz, Customer Intelligence Uniqa Insurance Australia	**Pavilion 4**
	Data Mining in Marketing **Determining the Best Balance of In-House and Outsource for your Segmentation and Building Internal Analytic Competence** Will Neafsey, Consumer Segmentation and Analytics Manager Ford Motor Company	**Tropical A-D**
	New and Emerging Technologies **Practical and Flexible Modeling with the GNBC: A Case Study** Kim Larsen, Director, Database and Relationship Marketing Charles Schwab & Co., Inc.	**Tropical E-H**
3:45 p.m.	**Break** *Refreshments served in the Exhibit Hall*	**Pavilion 7**
4:15 – 5:15 p.m.	**Breakout sessions**	
	Perspectives from Healthcare **Mining Medical Claims Data from Exploratory to Confirmatory Statistical Methods** Thomas T.H. Wan, Ph.D., M.H.S., Professor and Associate Dean for Research College of Health and Public Affairs University of Central Florida	**Palma ABEF**
	Business Applications **Using Data Mining to Build an Effective Recruitment and Retention Strategy** Cali M. Davis, Enrollment Research and Assessment Analyst J. Michael Hardin, Director of Business Intelligence and Professor of Statistics University of Alabama Tom Bohannon Baylor University	**Palma CDGH**
	Fraud Detection **Data Preparation for Insurance Fraud Detection** Terry Woodfield, Statistical Services Specialist SAS Institute Inc.	**Pavilion 3**
	Perspectives from Retail **What's Leaving Your Wallet? Estimating Industrial Share of Spending in a Competitive Environment with Incomplete Information** Talbot Michael Katz, Analytic Data Information Technology Consultant	**Pavilion 4**
	Data Mining in Marketing **"Supervised" Customer Segmentation Analysis using SAS® Enterprise Miner™** Paul B. (Brad) Jordan, Leader, Marketing Analytics Blue Cross Blue Shield of Florida	**Tropical A-D**
	New and Emerging Technologies **Linear Models, Colinearity and Variable Selection for Giga Bases** Leonardo Auslender, Statistician and Economist SAS Institute Inc.	**Tropical E-H**

Tuesday, October 24

7:30 a.m.	**Registration open** *Continental breakfast available in the Exhibit Hall*	Salsa Registration Pavilion 7
8:45 a.m.	**Welcome from Co-Chairs**	Pavilion 9
9:00 a.m.	**Keynote Address** **Elliptical Predictors in Logistic Regression** Will Potts, Vice President of Statistical Analysis Capital One	Pavilion 9
10:00 a.m.	**Break** *Refreshments served in the Exhibit Hall, sponsored by Capital One*	Pavilion 7
10:30 a.m.	**Keynote Address** **Bringing Data Mining Down to Business** Randy Collica, Senior Business Analyst Hewlett-Packard	Pavilion 9
11:30 a.m.	**Lunch** *Exhibit Hall open*	Amazon A
12:30 p.m.	**Keynote Address** **Business Intelligence Success Factors—Essential Skills for Success in a High-tech, Data Driven Corporate World** Olivia Parr-Rud, Consultant/Facilitator, OLIVIAGroup	Pavilion 9
1:30 – 2:30 p.m.	**Breakout sessions**	
	Perspectives from Healthcare **An Epidemiological Framework for a Hypothesis Generating Investigation of Multiple Near-Concurrent Vaccinations and Potential Health Endpoints** Daniel C. Payne, Ph.D., MSPH, Principal Epidemiologist and Project Officer U.S. Centers for Disease Control and Prevention	Palma ABEF
	Business Applications **Preparing the Data Mining Data: Techniques for Harmonizing and Integrating Information from Disparate Datasets** James (Jim) W. Mentele, Senior Research Fellow Central Michigan University, Research Center	Palma CDGH
	Fraud Detection **Operationally Significant Patterns of Association** William C. Hardy, Senior Metrics Analyst Richard W. La Valley, Senior Metrics Analyst Advanced Systems Concepts SAIC, Inc.	Pavilion 3
	Business Applications **Business Price Optimization using Data Mining and Modeling** Timothy D. Rey, Manager, Data Mining and Modeling The Dow Chemical Company	Pavilion 4
	Data Mining in Marketing **Analytical Roadmap—the Marketer's Map to Improved ROI** Maria Marsala Herlihy KnowledgeBase Marketing, Inc.	Tropical A-D
	New and Emerging Technologies **Exploring Open Source Software Development and Maintenance Using Data Mining and Text Mining and Modeling** Uzma Raja Dept. of Information Systems, Statistics, and Management Science, Culverhouse College of Commerce The University of Alabama	Tropical E-H

2:45–3:45 p.m.	**Breakout sessions**	
	Perspectives from Healthcare **Training Healthcare Professionals in Data Mining** Mark E. Johnson, Professor of Statistics Morgan Wang, Director, Institute of Statistics and Data Mining University of Central Florida	Palma ABEF
	Business Applications **Data Preparation for Analytics** Gerhard Svolba, Consultant SAS Institute Inc., Austria	Palma CDGH
	New and Emerging Technologies **New Mathematical Optimization Functionality in SAS/OR®** Manoj Chari, Manager, Marketing Optimization (R&D Group) SAS Institute Inc.	Pavilion 3
	Perspectives from Financial Services **Bayesian Methods in Asset Management and Risk Management** Joe Chen, Ph.D., Vice President, Risk Modeling Origen Financial	Pavilion 4
	Data Mining in Marketing **Leveraging Advanced Statistical Analysis to Turn Online Marketing Data into Actionable Marketing Insights** Jack H. Schwartz CTO, [x+1]	Tropical A-D
	New and Emerging Technologies **They Asked for a Segmentation Scheme, not Clusters** Jeff Zeanah, President Z Solutions, Inc.	Tropical E-H
3:45 p.m.	**Break** *Refreshments served in the Exhibit Hall*	Pavilion 7
4:15–5:15 p.m.	**Breakout sessions**	
	Perspectives from Healthcare **Data Mining to Determine Characteristics of High-cost Diabetics in a Medicaid Population** S. Greg Potts, MBA, Data Mining Team Leader Arkansas Foundation for Medical Care Office of Projects and Analysis	Palma ABEF
	Business Applications **Data Integration – Tools vs. Code. Tips and Techniques for Converting from User-written Code to Process Workflows Within SAS Data Integration** Cary White, Data Warehouse Team Leader University of North Carolina, Chapel Hill	Palma CDGH
	Data Mining in Education ***Abstract not available at printing*** James Ashby, Director of Assessment and Accountability Plano Independent School District, TX	Pavilion 3
	Perspectives from Financial Services **Financial Data Mining: What to do with the Money You Bring Home from Las Vegas** Gary D. Boetticher, Associate Professor University of Houston, Clear Lake	Pavilion 4
	Data Mining in Marketing **Data Mining—Interactive Marketing** Kim Bell Data Analysis, Inc.	Tropical A-D
	New and Emerging Technologies **Word of Mouth and Opinion Spreading: New Sociophysics Approaches** Dmitri Kuznetsov, Ph.D., Senior Brand Analyst Media Planning Group	Tropical E-H

Monday Keynote Speakers

9:00 a.m. – 10:00 a.m.
A View of the Data Mining Revolution and Evolution

William B. Smith
Executive Director,
American Statistical Association
and Professor Emeritus of Statistics, Texas A&M University

With the advent of modern and more powerful computers with inexpensive memory and automatic measuring and collecting devices and techniques, the volume (n) of multidimensional (p) data available for 'analysis' has increased beyond belief. As a statistician, the speaker will trace the revolution/evolution by illustrating procedures appropriate for various combinations of n and p. Traditional statistical methods focused on problems assuming small n and small p. Small n and large p problems (commonly called 'the curse of dimensionality') have been attacked, also. However, given the massive information now available, the data mining community addresses 'large n, any p' problems. Examples of analyses will be given to illustrate the methodologies of the n and p combinations.

10:30 a.m. – 11:30 a.m.
From Data Mining Grand Challenges to the New Sciences Underlying the Internet

Usama Fayyad
Yahoo! Inc.
Chief Data Officer and Senior Vice President of Research and Strategic Data Solutions

As the Internet continues to change the way we live, find information, communicate, and do business, it has also been taking on a dramatically increasing role in marketing and advertising. Unlike any prior mass medium, the Internet is a unique medium when it comes to interactivity and offers ability to target and program messaging at the individual level. Coupled with its uniqueness in the richness of the data that is available, in the variety of ways to utilize the data, and in the great dependence of effective marketing on applications that are heavily data-driven, makes data mining and statistical data analysis, modeling, and reporting an essential mission-critical part of running the business. However, because of its novelty and the scale of data sets involved, few companies have figured out how to properly make use of this data.

In this talk, I will review some of the challenges and opportunities in the utilization of data to drive this new generation of marketing systems. I will provide several examples of how data is utilized in critical ways to drive some of these capabilities. The discussion will be framed with the more general framework of Grand Challenges for data mining: pragmatic and technical.

I will conclude this presentation with a consideration of the larger issues surrounding the Internet as a technology that is ubiquitous in our lives, yet one where very little is understood, at the scientific level, in defining and understanding many of the basics the Internet enables: Community, Personalization, and the new Microeconomics of the web. This leads to an overview of the new Yahoo! Research and its aims: inventing the new sciences underlying what we do on the Internet, focusing on areas that have received little attention in the traditional academic circles.

12:30 p.m. – 1:30 p.m.
Building a Data Mining Community

David Duling
SAS Institute Inc.
Software Development Manager, SAS® Enterprise Miner™

Data miners typically work in a very competitive environment. Statistical models that impact business success are treated as corporate assets that are subject to tight security. This includes the business purpose, data selection, target definition, feature aggregation, and model technique. Data miners often acquire techniques from vendors and ideas from conferences, journals, and informal discussions. Enterprise Miner extension nodes provide a mechanism for sharing basic functions between users in an organization, between consultants and customers, and ultimately between users across organizations. This presentation discusses the extension node mechanism, presents examples of nodes that have proven to be useful, and proposes a program for exchange between users.

Monday Breakout Sessions 1:30 – 2:30 p.m.

Mining Without a Hardhat: A New Approach to Safely at a National Laboratory

Judy Spomer
Knowledge Discovery and Extraction,
Sandia National Laboratories

Sandia is a national security laboratory involved in a variety of research and development programs to help secure a peaceful and free world through technology. We develop technologies to sustain, modernize, and protect our nuclear arsenal, prevent the spread of weapons of mass destruction, defend against terrorism, protect our national infrastructures, ensure stable energy and water supplies, and provide new capabilities to our armed forces.

The widespread nature of activities conducted at Sandia presents a unique challenge: How to keep approximately 8,500 staff members safe and injury-free when the work involves radiological materials, explosives, chemicals, lasers, and so on?

In early 2004, the Environment, Safety, and Health organization at Sandia National Laboratories initiated an effort to develop a statistical model to identify characteristics that could forecast the likelihood of work-related injuries or illnesses by organization. The resulting Injury and Illness Predictive Model (IIPM) has been in production since December 2004. On a quarterly basis, current data is run through the model, generating a forecasted 6-month safety incident rate for organizations at the laboratories. Resources can then be focused on interventions for areas at risk.

This session will cover the approach, stumbling blocks, and methods used to develop this model in SAS as well as its deployment at Sandia.

Why Data Mining Is an Encapsulating Solution to Querying, BI and BA in Research and Planning in Higher Education and Beyond

Jing Luan, Vice Chancellor,
Educational Services and Planning,
San Mateo CCCD

The session covers the general use of data mining applications for higher education research and decision making with a focus on understanding data mining being the ultimate system approach to informatics. Specifically, since data mining resides on the top of the 3 tiered Knowledge Management model developed by Dr. Jing Luan, he will describe the use data mining applications as an end game for any of the three main purposes of data mining: data queries, report generation (BI), and predictive modeling (BA).

Needs Based Segmentation and its Application in SKTelcom

Jongdo Kim, SKTelcom, Korea

As services and products grow various and complicated, it is harder for customers to find right service for themselves and for company to find right target. In this session, how SKTelecom is segmenting customers and using segmentation to offer services will be discussed. We segmented customers based on their needs, that we assume to be revealed by their behaviors. We are using this segmentation to increase sales and to enhance our customer satisfaction.

Detecting Fraud by Building Strong Control Environments using Continuous Monitoring

William (Bill) J.Kelley, COL FA, Program Director, Data Mining Division, Department of Defense, Office of the Inspector General

Some topics to cover include:

* Establish effective review and approval processes* Utilize human capital management to foster accountability

* Implement innovative monitoring procedures to eliminate fraud and abuse by applying business rules that identify transactions with high risk

JUSTIFICATION:
With the continued downsizing of oversight resources—governmental agencies and the private sector need to take advantage of continuous monitoring techniques to identify transactions that have high risk. Although the DoD has taken aggressive action to improve its programs, most programs still need better implementation and oversight of management controls at the corporate and activity level.

Using SAS® Data Mining Procedures to estimate the 'true' value of new (and existing) products and features

Abdu Elnagheeb, Senior Vice President, Bank of America

One of the more difficult analytic challenges faced by financial and other industries is to determine the incremental profitability of products that are rushed to market without the benefit of a well thought out experimental design. It is not uncommon for a product or feature to be rolled out across the franchise, and then have senior management ask: "How much incremental profit is being generated?"

Too often product managers will report that customers buying the new product are worth 'X', customers buying the old product are worth 'Y', so the incremental value of the product is X - Y. But common sense and experience tell us that the customers selecting X over Y maybe be fundamentally different customers, and that self-selection, differential opportunity, and other inherent biases invalidate such analyses.

To overcome the influence of these biases, we have developed a methodology that uses CHAID, Linear, and other modeling techniques to back out customer and market influences, and more precisely zero in on the true value of the product.

This presentation follows the development of the methodology using artificial as well real data and actual products to demonstrate how the true value of a product can be estimated, even when there are strong biases inherent in the purchase and use of the product. Variable selection, model optimization, actual code, and the shortcomings of the mythology are presented.

Predictive Modeling and Segmentation in the Newspaper Industry

Goutam Chakraborty, Ph.D., Associate Professor of Marketing, Oklahoma State University

In this talk, I will discuss an application of predictive modeling and segmentation for a major newspaper in the southwestern United States. The data environment includes both address-specific (individual household level) data as well as grouped (census block group level) data. Modeling task includes predicting subscription (yes/no) as well as the type of subscription (Regular including Sunday, Regular but no Sunday, 3-day weekend and Sunday-only) for households.

Visualizing Multiple and Logistic Regression Models

George Fernandez, Professor of Applied Statistics, University of Nevada, Reno

Data mining techniques stress visualization to study thoroughly the structure of data and to check the validity of statistical model fit to the data. Multiple and binary logistic regression are widely used supervised learning methods in data mining. In multiple linear regression models, problems arise when serious multicollinearity or influential outliers are present in the data. Failure to include significant quadratic or cross-product terms results in model specification error. Simple scatter plots and logit plots are not effective in revealing the complex relationships of predictor variables or data problems in multiple linear and in logistic regression. In multiple linear regression applications, partial regression plots are considered useful in detecting influential observations and multiple outliers; partial residual plots or the added-variable or component-plus-residual plots are useful in detecting non-linearity and model specification errors. The leverage plots are considered effective in detecting multicollinearity and outliers. The VIF-plot, which is very effective in detecting multicollinearity, can be obtained by overlaying both partial regression and partial residual plots with a common centered X-axis. In case of binary logistic regression (BLR), the partial delta logit plots are useful in detecting, significant predictors, non-linearity, and multicollinearity. The partial delta logit plot illustrate the effects of a given continuous predictor variable after adjusting for all other predictor variables on the change in the logit estimate when the variable in question is dropped from the BLR. By overlaying the simple logit and partial delta logit plots, many features of the BLR could be revealed. User-friendly SAS macro applications for generating these exploratory plots will be presented.

Using Administrative Records to Assess the Performance of Health Care Providers— Pitfalls and Challenges

HungChing Chan, Director of Clinical Analysis and Program Evaluation, Medica

Health insurance claims capture the majority of health services transactions, and provide a potential data source to improve quality and reduce costs of health care. Several case studies will be presented to demonstrate how insurance claims are used for these purposes. The challenges of mining health insurance transaction data will be discussed as well.

Customer Acquisition and Retention using Data Mining Techniques

T. Lynn Locke, Director, Database Marketing, Blue Cross Blue Shield of Florida

A complete case study that illustrates how to utilize data mining techniques to enhance the customer acquisition process will be presented. The resulting model from this study has been tested in January and April 2006. Utilizing this model, the marketing department can afford to mail four times more volume than in prior years, while reducing the per customer acquisition cost by 67%.

Data Mining at the Texas State Auditor's Office

Tom Winn, Senior Systems Analyst, Texas State Auditor's Office

In general, data mining is the iterative and interactive process of finding new and potentially useful knowledge from large quantities of data. The author's use of that term encompasses sophisticated statistical techniques, as well as many simpler data analysis methods. The mission of the Texas State Auditor's Office is to actively provide government leaders with useful information that improves accountability. This expository presentation will describe some data mining work at the Texas State Auditor's office in support of auditing and investigations. Particular attention will be given to fraud detection and risk assessment in the public sector.

Perspectives from Retail
Location: Pavilion 4

Customer Profit Value in Insurance Business

Günter Schmölz, Uniqa Insurance Austria, Customer Intelligence

UNIQA is Austria's biggest insurance company, is active in more than 14 European countries and has more than 6 million customers. Uniqa has built up an international, comprehensive and standardized SAS-database architecture which enables an integrated data management process. This database is not only the basis for the data mining process with SAS Enterprise Miner but also the central basis for the whole campaign management, for the Client Reporting system, for all actuarial analyses and an important source for the SAS Balanced Scorecard.

This presentation shows how data mining and customer-data-management triggers Uniqa's client orientated strategy. The main results (individual client ratios) of the data mining and scoring process are implemented in the UNIQA client information system (which all employees have access to). This presentation will concentrate on the most

important ratio, the Customer-Profit-Value. This Value is a prognosis of the expected client contribution margin for the next 12 months. The presentation will give an insight on how more than 50 customer-criteria are influencing the individual score. The Customer-Profit-Value enables UNIQA to differentiate between good and bad customers and allows a very precise value prognosis not only on an individual basis but also on a segment level. This presentation will also show how UNIQA uses these customer orientated scores to steer the sales force.

Data Mining in Marketing
Location: Tropical A-D

Determining the Best Balance of In-House and Outsource for your Segmentation and Building Internal Analytic Competency

Will Neafsey, Consumer Segmentation and Analysis Manager, Ford Motor Company

As market research and brand organizations are constantly streamlined and rationalize with today's business environment, marketers are continually faced with the decision of outsourcing their analysis or doing it in-house. This can be a complex decision, particularly with segmentation. The application of analytic and segmentation techniques can be very industry specific. Valuable time and money can be wasted educating a consultant on your industry, your company and your consumers. On the flip side, hiring, training, staffing and protecting an analytic/segmentation team can be difficult. This session will discuss the advantages and pitfalls of bringing segmentation and analytic competency in-house

New and Emerging Technologies
Location: Tropical E-H

Practical and Flexible Modeling with the GNBC: A Case Study

Kim Larsen, Director, Database & Relationship Marketing, Charles Schwab & Co., Inc.

Predicting the probability of a binary event given a set of independent variables is one

of the most commonly occurring modeling problems, whether it's risk management, marketing, or a medical study. Last year at M2005, I presented the Generalized Naïve Bayes Classifier (GNBC) and argued that it's one of the most powerful tools for task. In short, the GNBC is a generalization of the well-known Naïve Bayes Classifier that:

• Treats input variables in a flexible nonparametric fashion

• Uncovers hidden patterns in the data

• Can handle problems of large dimensions

This year I'll re-visit the GNBC, but spend less time on theoretical background and justification, and instead concentrate the talk around a case study using real-life data. The case study will demonstrate:

• How to use, interpret, and implement the GNBC

• The power of the GNBC and how it compares to other well-known techniques.

Monday Breakout Sessions 4:15 – 5:15 p.m.

Perspectives from Healthcare
Location: Palma ABEF

Mining Medical Claims Data from Exploratory to Confirmatory Statistical Methods

Thomas T.H. Wan, Ph.D., M.H.S., Professor and Associate Dean for Research, College of Health and Public Affairs, University of Central Florida

Health services researchers have relatively concentrated in applying exploratory statistical methods in identifying the patterns of care and analyzing the variation in health services use. Despite the critical role medical claims data play in detecting billing errors or fraudulent behaviors of billing practice, little is known about the extent to which variations in frequency, types, and seriousness of deficiencies reflect actual differences in quality of care versus broader systemic differences

in the utilization and provision of health services. In response, we propose to formulate an evidence-based approach to mining claims data for not only identifying patterns of billing errors, but also examining the relative influences of facility/organizational and contextual factors that may influence a specific pattern of Medicare/Medicaid fraud and abuse. Examples of applying exploratory and confirmatory statistical methods, using claims data, are presented.

Business Applications
Location: Palma CDGF

Using Data Mining to Build an Effective Recruitment and Retention Strategy

Cali M. Davis, Enrollment Research and Assessment Analyst,
The University of Alabama
J. Michael Hardin, Director of Business Intelligence and Professor of Statistics,
The University of Alabama
Tom Bohannon, Assistant Vice President and Director, Institute Research and Testing Services,
Baylor University

Freshmen recruitment, enrollment, and retention are important issues in higher education. Researchers from across the country have used data mining techniques to develop models to predict who will enroll and why students fail to enroll for their second year. Over the past several years, the University of Alabama, a public university, and Baylor University, a private university, have used SAS Enterprise Miner and data mining to develop predictive models from admissions and records data.

In this presentation, the data mining experiences of both of these universities will be discussed, as well as a comparison and contrast of the private and public university perspective. In addition, other applications of data mining in higher education such as donor modeling and student recruitment strategies will be discussed.

Fraud Detection
Location: Pavilion 3

Data Preparation for Insurance Fraud Detection

Terry Woodfield, Statistical Services Specialist, SAS

Detecting fraudulent claims in property and casualty insurance requires not only good data, but domain expertise in deriving inputs for fraud modeling. Many of the variables available from an insurance company's database are categorical, such as gender, ICD-9 codes, employment classifications, accident codes, and body part codes. Diagnostic codes like ICD-9 codes have very high cardinality and require either automatic or manual grouping. Strategies for converting categorical variables into predictive modeling inputs will be described. Converting transactional data, such as medical payments, into predictive modeling inputs will also be discussed.

Perspectives from Retail
Location: Pavilion 4

What's Leaving Your Wallet? Estimating Industrial Share of Spending in a Competitive Environment with Incomplete Information

Talbot Michael Katz, Analytic Data Information Technology Consultant

A holistic view of a customer base is a powerful weapon in the hunt for consumer dollars, but we typically confront several holes in our efforts to construct such a portrait. We are almost always missing a great deal of information about the full spending profile of each customer, and the customers for which we have the best data may not be representative of the entire base. And although spending in separate industries may be influenced by unrelated factors and have very different characteristics, because all dollars are in competition, it may be desirable to treat the entire spectrum of industries in concert, especially to be able to consider the key element of elasticity. We will examine

how to piece together a picture of wallet share using classic methodologies such as segmentation, discrete choice modeling, and seemingly unrelated regressions, keeping in mind such issues as handling of missing data and variable selection. And we will look at promising new approaches to explicitly exploit the gaps in information, such as Chen and Steckel's hierarchical Bayesian analysis of interpurchase times.

Data Mining in Marketing
Location: Tropical A-D

"Supervised" Customer Segmentation Analysis using SAS Enterprise Miner

Paul B. (Brad) Jordan, Leader, Marketing Analytics
Blue Cross Blue Shield of Florida

Customer segmentation is one important way to increase marketing efficiency. Traditionally, clustering algorithms are used this goal. However, clustering techniques are unsupervised and there is not any easy way to assess the clustering performance. Using a "supervised" data mining approach has helped us achieve the desired result of producing segments that are linked to a target variable.

New and Emerging Technologies
Tropical E-H

Linear models, Colinearity and Variable Selection for Giga Bases

Leonardo Auslender, Statistician and Economist, SAS

Variable selection is heavily used in the process of model building with large giga bases of observational data. While variable selection methods, typically of the stepwise family, are (or should be) fairly well understood, issues such as multicolinearity become the bete noir especially when the resulting model search it not very successful. We analyze the merits of these claims, and propose a user-controlled tool that remedies colinearity resulting of the variable selection search.

Announcing the Data Mining Shootout
presented by SAS and Inductis

Win one of three prizes ($6,000, first prize; $4,000, second prize; $2,000, third prize), a free trip to M2007 and the opportunity to present your work at the largest data mining conference in the world!

In 2007 SAS and Inductis will co-sponsor the first annual Data Mining Shootout modeling competition, open to academia. This contest will provide a problem and identical testing data to contestants, asking them to build a model using SAS software to solve the problem. The three most impressive predictive modeling submissions will be awarded with an all-inclusive trip to Las Vegas to attend M2007 and prize money equaling the submission's finish: $6,000, first prize; $4,000, second prize; $2,000 third prize. Each of the top three submissions will also be required to present their results in a 15 minute presentation at M2007.

Each monetary award is divided equally between a cash scholarship (to be presented to the winning institution's departmental contact) and a cash payment made to the team lead. Each of the top three submissions will also receive airfare, hotel, meals and a free conference registration for the project lead. In order to compete, you must be a full-time student, staff or faculty member at an accredited university or college. Registration begins October 23, 2006 (at M2006) and continues until February 15, 2007. Registrations must be made through the online registration form at **www.inductis.com**. (Each team will receive the competition's testing data at the same time – on or shortly after February 15.) A project lead must be named on each submission form.

All submissions must be complete and delivered by June 30, 2007. Submissions will be judged by a review committee selected by SAS and Inductis and a decision will be made by July 31, 2007. For more information e-mail **m2007@inductis.com**.

§sas | **THE POWER TO KNOW.**

INDUCTIS
IDEAS ▸ ANALYSIS ▸ EXECUTION

Visit the Inductis booth in the Exhibit Hall to enter the competition.

Tuesday Keynote Speakers

9:00 a.m. – 10:00 a.m.
Elliptical Predictors in Logistic Regression

Will Potts
Capital One
Vice President of Statistical Analysis

Many customer attributes have nonnormal features such as skewness and heavy tails. Logistic regression does not involve distributional assumptions on the predictors. Nevertheless, normalizing transformations can make it easier to discover models that are both simple and powerful. An elliptically contoured predictor space sharpens the relationship between the response and explanatory variables by reducing nonlinear confounding and the influence of sparse regions. The benefits of predictor transformations are demonstrated using credit bureau data.

10:30 a.m. – 11:30 a.m.
Bringing Data Mining Down to Business

Randy Collica
Hewlett-Packard
Chief Data Officer and Senior Vice President of Strategic Data Solutions

Anyone who has performed or been involved with data mining and business intelligence will probably recognize that one of the first questions to be answered is: "What is the main business problem or issue that needs to be solved?" Too often management's expectation of data mining is "help me find something interesting" or "what can you do with my data." In all of these situations, translating the business problem or issue into a form that can be addressed by data mining or modeling is the first order of business. Then, setting expectations for all aspects of the project will go a long way at ensuring that there are no real "surprises" in the results. Understanding the main business problem should precede any data mining or modeling and having this understanding at the outset will increase the likelihood of success in the overall project.

This presentation will look at several case studies in data mining and business intelligence projects and their true business issue or pain. Key findings and experiences will be shared with examples in survival data mining, forecasting, structured data mining, and text mining. Another key learning that will be addressed is the how to explain results from complex models and analyses to an audience that has little experience with data, statistics, or modeling.

12:30 p.m. – 1:30 p.m.
Business Intelligence Success Factors—Essential Skills for Success in a High-tech, Data-driven Corporate World

Olivia Parr-Rud, OLIVIAGroup
Consultant/Facilitator

Data mining and business intelligence solutions offer incredible opportunities for companies to improve their bottom line. However, increased complexity and the constant introduction of new technologies as well as shifting markets and consumer demands require us to master a broader set of skills. Expert leadership has never been more important. Studies show that roughly 70% of business intelligence projects fail due to poor leadership. This presentation will discuss concepts and provide insights into some basic skills that are essential for success in a high-tech, data-driven industry that is quickly going digital, virtual and global.

Communication is an essential skill in the complex world of business intelligence. As we turn to high tech solutions for our database marketing challenges, our employees require highly specialized and varied skill sets. In many organizations, each group speaks a "different language." Yet their ability to effectively communicate is critical to our business success.

Olivia will be signing her book, Data Mining Cookbook *during the Monday evening reception at the Bookstore inn the Exhibit Hall.*

Tuesday
Breakout Sessions
1:30 – 2:30 p.m.

Perspectives from Healthcare
Location: Palma ABEF

An Epidemiological Framework for a Hypothesis Generating Investigation of Multiple Near–Concurrent Vaccinations and Potential Health Endpoints

Daniel C. Payne, PhD, MSPH[1], Principal Epidemiologist and Project Officer, U.S. Centers for Disease Control and Prevention

Additional Authors:
Charles E. Rose, PhD[1], Aaron Aranas, MPH, MBA[1], Michael M. McNeil, MD, MPH[1], Jodi Blomberg, MS[2]

Affiliations: [1] U.S. Centers for Disease Control and Prevention, National Immunization Program, Epidemiology and Surveillance Branch, Bacterial Vaccine–Preventable Diseases Branch, [2] SAS Institute.

Limited scientific knowledge exists regarding whether health effects are associated with receiving multiple near–concurrent (MNC) vaccinations among adults. The medical literature available on this topic addresses primarily isolated observations following specific childhood vaccine combinations. To our knowledge, no defined framework for evaluating this topic in a comprehensive and standardized fashion using a large health outcomes dataset has been reported. A prototype framework using SAS Enterprise Miner applications to study this topic will be presented.

Staff from the Centers for Disease Control and Prevention (CDC) and SAS, Inc. applied SAS Enterprise Miner software to data collected by the Defense Medical Surveillance System (DMSS), a longitudinal surveillance database administered by the Army Medical Surveillance Activity. We used SAS Enterprise Miner to conduct a range of analyses, including decision tree modeling and sequence analysis, to define and describe this hypothesis generating framework using the full DMSS dataset of medical, vaccination, administrative, demographic, and deployment information for over 7 million individuals. Our approach enabled analysis of the topic of MNC vaccinations and health endpoints to be viewed from 4 relevant perspectives of exposure: differing combinations of nominal vaccine types, live versus non-live vaccines, vaccines containing aluminum adjuvant versus those without this adjuvant, and lastly, by a continuous variable representing cumulative antigen counts.

Business Applications
Palma CDGH

Preparing the Data Mining data: Techniques for harmonizing and integrating information from disparate datasets

James (Jim) W. Mentele, Senior Research Fellow, Central Michigan University—Research Center

This talk will focus on the activities of data mining projects typically requiring 80-90% of the project effort. Harmonizing and integrating data from multiple functions and departments, customers & vendors, governments and 3rd party data sources, over various time-periods and levels of granularity presents challenges. The author will discuss an overview of techniques with examples from 40 years of related experience.

Fraud Detection
Location: Pavilion 3

Operationally Significant Patterns of Association

William C. Hardy,
Senior Metrics Analyst
Richard W. La Valley,
Senior Metrics Analyst
SAIC, Inc.
Advanced Systems Concepts

An association between two persons is an observation or reported fact that can reasonably be assumed to indicate they know of, and probably communicate with, each other. Much of the information on known/suspected terrorists comprises collections of such associations absent information that would support determination of whether any particular set of associations evinces cooperative involvement in pursuit of a common goal or objective. A major challenge for information technology is to develop tools for discovering such stronger relationships in large collections of associations gathered from disparate sources. To this end, this paper describes patterns of associations that are indicators of possible organizational activity, regardless of the means of their discovery. It describes for each:

- techniques for visualizing the pattern in the data;

- efficient means of discovery of instances of the pattern in very large link data bases; and

- procedures for validating instances of the pattern as being attributable to organizational behavior rather than a chance encounters.

It further describes results of analyses of real-world data that demonstrate the operational utility of inferences derived from the patterns.

Business Applications
Location: Pavilion 4

Business Price Optimization using Data Mining and Modeling

Timothy D. Rey, Manager, Data Mining and Modeling
The Dow Chemical Company

Businesses at Dow are always trying to set their prices correctly to obtain a reasonable margin for operating the business. This is a complex problem in that many issues can come into play: Raw Material Costs, Plant Upsets, Accounting Practices, Supply/Demand in the market, Customer buying patterns, to name a few. This project combined many data sets to study internal and external trends. SAS Enterprise Miner was initially used to the do the exploratory Data Mining. Eventually an iThink nonlinear simulation model was build for the business to use. The Dow Data Mining and Modeling process followed to conduct the project will be reviewed along with the interim results and final model.

Analytical Roadmap – the Marketer's Map to Improved ROI

Maria Marsala Herlihy, KnowledgeBase Marketing, Inc.

If you're ready to take your marketing campaigns to the next level, this session is for you! Many marketers want to employ analytics but don't know where to start. This session introduces you to the Analytical Roadmap AND you won't need a statistics degree to comprehend. All the concepts are presented in terms that you will understand illustrated with real-life case studies from companies that benefited from using analytical tools. This is a must-attend session for marketing professionals looking to improve their ROI!

You'll learn:

• Complex analytical concepts in familiar terms that you can understand.

• When to use analytics, how to get started and which tools are appropriate for your marketing challenges.

• How to bridge the gap between traditional direct marketing and database marketing

• Practical examples of how to apply analytics for increased ROI.

New and Emerging Technologies
Location: Tropical E-H

Exploring Open Source Software Development and Maintenance Using Data Mining and Text Mining

Uzma Raja, Department of Information Systems, Statistics, and Management Science, Culverhouse College of Commerce, The University of Alabama

Experiences of an exploratory study of the Open Source Software development and maintenance will be shared in this talk. The corporate use of OSS is increasing and there

is a need for evaluation models of these projects. In this research data mining and text mining techniques were used to discover knowledge from transactional datasets maintained on OSS projects. Models were formulated, tested and validated using these very large and comprehensive datasets. SAS Enterprise Miner and SAS Text Miner were used to develop the performance evaluation models.

In this presentation, the data mining and text mining results of the OSS projects will be discussed. The various factors that affect the outcomes of OSS projects will be highlighted. In addition some experiences using text mining to improve model performance will also be discussed.

Tuesday Breakout Sessions 2:45 – 3:45 p.m.

Perspectives from Healthcare
Location: Palma ABEF

Training Health Care Professionals in Data Mining

*Mark E. Johnson, Professor of Statistics
Morgan Wang, Director, Institute of Statistics and Data Mining
University of Central Florida*

Since M2005, we completed the delivery of a 5 session (3 days per session) data mining short course for a major health care insurance company. This course was primarily geared for marketing and information technology professionals who had considerable experience with SAS but limited exposure to SAS Enterprise Miner. Highlights of the training will be described including the incorporation of company specific projects and carefully selected pedagogical data sets. Guidelines on the balance among theory, application and output interpretation will be given. Finally, comparisons and contrasts of this data mining training experience to other training in which the presenters have been

involved (data mining, black belt six sigma training and design for six sigma) will be provided.

Business Applications
Location: Palma CDGH

Data Preparation for Analytics

*Gerhard Svolba, Consultant
SAS Institute, Austria*

Data preparation for analytics is an essential task for successful analysis, data mining and predictions. The quality of the data mart and its preparation is one of the most important influence factors for model quality and analysis result.

Data preparation itself must not only be considered as a technical or coding task. Analytic data preparation includes also the business point of view of the underlying business question as well as analytical and data modeling considerations. The business question for example may impose restrictions on which periods may be used as input data. Different analytical approaches require different analytic data structures.

Nonetheless coding of a data mart is a very important task in data preparation including data extraction from source systems, table joins, aggregations and the creation of derived variables. Time saving and clever approaches for data preparation facilitate the data preparation phase and provide derived variables with high explanatory and predictive power.

The presentation will deal with a specific case study and show how data from various data sources can be made up to build a clever one-row-per-subject data mart.

This presentation shows mainly SAS Code from SAS Base (Datastep, Procedures, SAS Macro Language) further a few procedures from SAS/STAT and SAS/ETS are used.

New Mathematical Optimization Functionality in SAS/OR

Manoj Chari, Manager, Marketing Optimization R&D Group, SAS Institute

Mathematical optimization is an analytical methodology that is closely associated with statistics and data mining—many predictive modeling and estimation procedures involve solving optimization problems, and many practical optimization based decision problems are based on input derived from statistical and data mining methods. In this presentation, we will discuss a completely new set of mathematical optimization procedures and solvers in SAS/OR whose initial release will be completed with SAS 9.2.

These include:

- PROC OPTLP for linear programming

- PROC OPTQP for quadratic programming with linear constraints

- PROC OPTMILP for mixed integer programming

- Two general nonlinear programming solvers based on sequential quadratic programming and interior point algorithms

- PROC OPTMODEL, a powerful algebraic modeling language with internal access to all of the above solvers.

Collectively these new procedures represent a quantum leap in functionality, performance, flexibility, and ease of use for SAS/OR optimization.

Using business and statistically oriented examples, we will present a unified view of this new optimization functionality through PROC OPTMODEL. This procedure combines a high level, flexible programming framework with algebraic modeling capabilities and an optimization oriented syntax. We will show how the mathematical formulation of an optimization problem that a modeler writes down

on paper can be passed virtually unchanged into OPTMODEL code. The procedure has an executable "solve" statement which can be used to invoke any of the above solvers. It is also possible, using the programming statements in PROC OPTMODEL, to implement customized optimization algorithms and heuristics that involve the iterative use of one or more types of optimization solvers.

Perspectives from Financial Services
Location: Pavilion 4

Bayesian Methods in Asset Management and Risk Management

*Joe Chen, Ph.D., Vice President
Risk Modeling, Origen Financial*

This presentation will include two parts:

Part 1
Black-Litterman model is widely applied by practitioners in sophisticated asset allocations. The model uses a Bayesian approach to combine investor's subjective beliefs on the expected returns of assets with the market equilibrium returns (the prior). Bayesian analysis provides a mechanism to synthesize subjective views with empirical realities. The Black-Litterman model resolves major issues of Markowitz's classical mean-variance paradigm. The latter is very sensitive to the covariance matrix and error-maximizing, and often ends up with unintuitive, highly-concentrated portfolios. Based on the posterior distribution of the new returns, the Black-Litterman model leads to intuitive portfolios with reasonable portfolio weights. Extension can be made if the normality assumptions are relaxed in the model.

Part 2
Under Basel II, banks can use their default probability estimates for calculating regulatory capital. Financial institutions use credit scoring models to underwrite and price loans. In either case, Bayesian inference can improve the accuracy of the default probability estimates. Bayesian methods also provide a natural way to handle structural differences between a bank's internal data and external data.

Data Mining in Marketing
Location: Tropical A-D

Leveraging advanced statistical analysis to turn online marking data into actionable marketing insights

Jack Schwartz, CTO, [x+1]

Today, the online customer touch point is the life's blood for the bulk of consumer-focused companies. In order to maximize marketing performance it is imperative that the most robust, consumer centric analytics be applied to campaigns to ensure the optimal experience for existing customers, as well as prospective ones.

When users move through websites they leave imprints which tells you a little about themselves, their preferences on a particular product or service and willingness to buy. Tracking their movement allows a marketer to better tailor their campaign to get this user to the ultimate goal – on-line purchase or repeat purchase.

But while the seasoned on-line marketer knows that this data will give them a competitive advantage, the primary challenge with web-analytics is to get them to act on the data. Now, as a growing number of companies return to data-driven marketing, some key techniques and principles need to be incorporated at the start of the online campaign to ensure solid results and sustainable growth.

New and Emerging Technologies
Location: Tropical E-H

They asked for a Segmentation scheme, not clusters

*Jeff Zeanah, President
Z Solutions, Inc.*

A common business request is to develop a segmentation scheme to assist in understanding customers and thus approach the market. The qualitative analyst generally converts these requests to some form of clustering using available data, because this is what

the analyst knows how to do. The restrictions of clustering approaches may have little to do with the actual business request leading to frustration for all parties. Using Genetic Algorithm Approaches in PROC GA in SAS/OR (experimental in 9.1.3, with expected release in 9.2) the analyst may optimize a segmentation scheme on any need that can be quantified. This presentation presents a simple application to introduce the concepts of Genetic Algorithms and initial approaches to accomplish an implementation in PROC GA.

Tuesday Breakout Sessions 4:15 – 5:15 p.m.

Perspectives from Healthcare
Location: Palma ABEF

Data Mining to Determine Characteristics of High-Cost Diabetics in a Medicaid Population
S. Greg Potts, MBA, Data Mining Team Leader, Arkansas Foundation for Medical Care, Office of Projects and Analysis

Diabetes Mellitus is a chronic condition affecting thousands of Arkansans. Arkansas Medicaid recipients with this condition often incur significant medical costs related to treatment of the condition.

Our objectives were twofold: 1) to use the Decision Tree algorithm in SAS Enterprise Miner 4.3 to determine high cost-drivers among Arkansas Medicaid recipients with Diabetes, and, 2) to determine whether the study cohort could be segmented into small, manageable groups for potential disease-management intervention in an effort to manage costs and provide quality care. Ours was a retrospective cohort study using all paid claims for recipients enrolled 12 continuous months during State Fiscal Year 2004 with a diagnosis of Diabetes Mellitus from Arkansas Medicaid's Decision Support System (DSS) Data Warehouse.

Business Applications
Location: Palma CDGH

Data Integration – Tools vs. Code. Tips and techniques for converting from user-written code to process workflows within SAS Data Integration.
Cary White, Data Warehouse Team Leader, UNC-Chapel Hill

Many shops have a large investment in developer-written ETL code. In many cases, this code was written by a consultant or developer who is no longer employed by the company. When a decision is made to purchase an ETL tool, the company must also decide how much of this code can be reused within the new tool.

The SAS language is a especially powerful 4 GL language that is well-suited for data extraction, transformation and loading (ETL) In its V9 release SAS is now offering an industry-standard ETL tool now called SAS Data Integration.

This talk will present a case study of a conversion project in which ETL processes written in Base SAS were converted to process workflows using SAS Data Integration Studio.

Some of the topics covered will include:

• Best practices to apply in the conversion process

• What to keep vs. what to convert

• Relationship of data quality to the data integration process

• Creating reusable processes

• Maintenance and Metadata—two of the most important reasons to use a tool

• Good features/shortcomings of the tool

Data Mining in Education
Location: Pavilion 3

Abstract not available at printing
James Ashby, Director of Assessment and Accountability, Plano Independent School District, TX

Perspectives from Financial Services
Location: Pavilion 4

Financial Data Mining: What to do with the money you bring home from Vegas
Gary D. Boetticher, Associate Professor, University of Houston – Clear Lake

Financial data mining has been described as "the toughest way to make easy money." There are many traps and misconceptions about financial modeling. Commercial products exist which exaggerate their findings or make faulty assumptions regarding their modeling process. Academic research may forget to discount "buy and hold" strategies in their results.

This session examines how to apply genetic programs to the stock market and optimize financial model building.

• It describes various traps in financial data mining process and offers different techniques for building sound financial models.

• It examines how to extend genetic programs for financial data mining.

• It describes different aspects of financial domain knowledge and how to incorporate it into the financial data mining process.

• It offers demonstrations on building financial models using stock data and/or future contracts.

Location: Tropical A-D

Data Mining — Interactive Marketing

Kim Bell, Data Analysis, Inc.

In the high-speed world of internet advertising, business owners are realizing that one of the most effective means of increasing online market share is through a combination of search engine optimization and pay-per-click advertising. Finding a way to help customers find you is the key to successful interactive marketing campaigns. Through data mining techniques, analyses are conducted on such categories as: keyword searches, visitor loyalty, visitor recency, search engine sources, website visits, page views, geographical locations, cost per clicks, advertised words (adwords), adword positions, and conversions from website visits to revenue dollars. Information used in the analyses includes, but is not limited to,

- the total number of visits and page views your site received

- the average number of page views per visit (P/V)

- the number of visits and page views over time (Averages are calculated over the entire selected date range including dates not yet elapsed when applicable)

- the number of first-time visits and returning visits

- the cities from which the most visitors come to your site

- top referral sources

- impressions (the number of times an ad was shown)

- clicks (the number of times an ad was clicked)

- trans (the number of transactions that occurred after a referral from this Source [Medium] or keyword)

- cost for internet advertising

This session will discuss the means of collaborating each of the above data points to increase website traffic, generate new business, lower customer acquisition costs, and enhance product awareness throughout the online marketplace.

Location: Tropical E-H

## Word of Mouth and Opinion Spreading: New Sociophysics Approaches

Dmitri Kuznetsov, Ph.D., Senior Brand Analyst, Media Planning Group

Understanding of leading mechanisms of opinion spreading in social systems is a key point for powerful marketing and political campaigns. Opinion exchange and convincing (so-called, "word of mouth") is a crucial part of the opinion spreading along with advertising, experience, and economic and political factors. During last decade opinion dynamics became one of the central parts of the scientific branch named Sociophysics that applies modern methods of statistical physics to social phenomena. Its success originates from the deep similarities in rising (1) from single-particle properties to complex-object features in physics and (2) from single person characteristics to social mass behavior. In 2005 we introduced Mediaphysics as a part of Sociophysics, studying processes of mass communications in social systems and demonstrated its potential for applications in different areas. In this presentation, different sociophysics approaches to the "word of mouth" and opinion spreading are demonstrated and discussed.

M2007

Knowledge that makes a **difference**

10th Annual
Data Mining
Conference
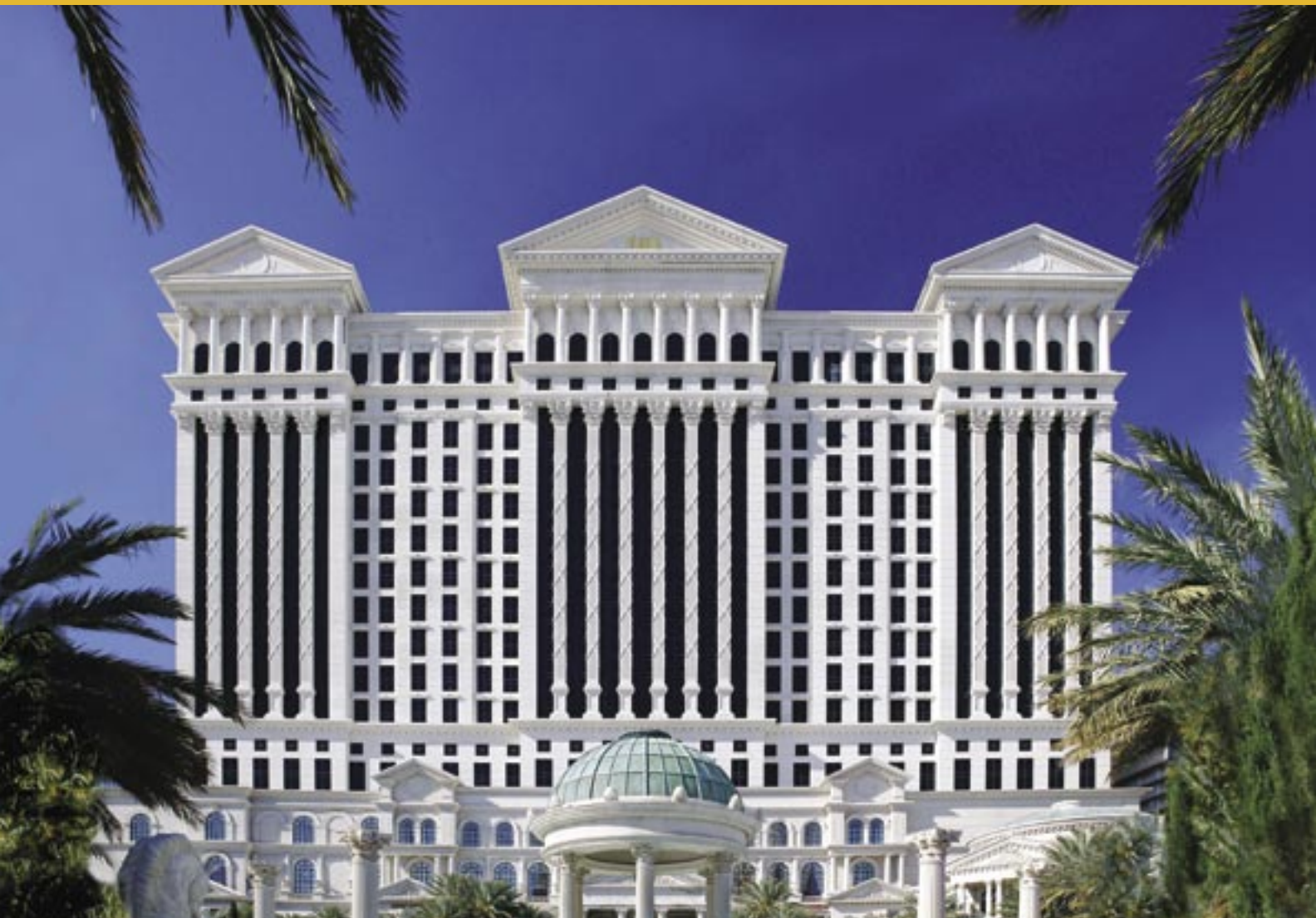
Come see us at our new location next year!

Caesars Palace
October 1-2, 2007

How the Doughboy graces millions of dinner tables. Always in a timely fashion.

Each day, Pillsbury products and other General Mills brands appear in millions of shopping carts around the world. HP Integrity servers with Intel® Itanium® 2 processors help keep distribution and inventory control systems running smoothly. "With their continuous performance and support, we are able to ensure that customer orders and shipments are processed quickly and accurately," said Vandy Johnson, Director of I.S. Operations. "And that's a comforting thought." itanium-integrity.com

ITANIUM + INTEGRITY. ON AND ON AND ON.

Meet the Speakers

James Ashby is the Director of Assessment and Accountability for Plano Independent School District, a Texas school system that has faced the challenges of educating an increasingly diverse population of students who are challenged in meeting state-mandated expectations. In Plano, James leads a team of research scientists focused on developing school leaders into relevant data-driven decision-makers.

Since 1997, James has worked closely with public school administrative teams assisting them in identifying systemic practices and policies that act as barriers to student achievement and equitable educational opportunities. Using statistical modeling that integrates national, state, and local assessment data with demographic and efficacy variables, James has empowered Texas administrators to make effective decisions that close significant gaps in student achievement and educational equity among diverse populations.

James is currently pursing a Ph.D. at the University of North Texas in Educational Statistics. His most recent published articles discuss the use of multivariate logistic regression for measuring teacher effects on student's cognitive growth.

Leonardo Auslender is a statistician and economist with more than 25 years of business experience and SAS® expertise, at present in the Data Mining Research and Development group of SAS Institute. His area of expertise is in the area of Giga-Data Analysis and Methods, and has written papers and given lectures on Missing Value Imputation, Classification Trees, Support Vector Machines, Market-Basket Analysis, Variable Selection in Giga-Bases, Database Marketing, CRM, GDP and (Relative Price) Inflation studies, Expectation Formations, Productivity and Technology effects in the economy, and most recently on Colinearity and malaise in linear modeling. He was a lecturer of Finance and Macroeconomics at Rutgers University. His present interests are in the area of variable selection, Bayesian networks, and Bayesian and Tree methods.

Kim W. Bell is president and founder of DataAnalysis, Inc., a worldwide provider of analytical and instructional services in the field of quality management. Her experience includes working with businesses in such diverse areas as the textile, chemical, government, automotive, and transportation and logistics industries, as well as in the healthcare, education, financial services, insurance, and research technology industries.

She has helped transition numerous companies to the most current quality improvement techniques, including the ISO 9001:2000 quality standard, the ISO 14001 environmental standard, Lean Manufacturing for the Manufacturing Floor, Lean Manufacturing for Office, Six Sigma, and Customer Relationship Management.

knowledge that makes a difference

In addition to providing customized on-site training and multi-company seminars, Kim has succeeded in developing unique solutions that enable multi-site corporations to collect and analyze their internal data. She also enjoys her partnership with SAS, which enables her to teach SAS statistical and quality improvement courses on a contractual basis.

Kim has chaired and served on numerous ISO conference committees. She is a member of the American Society for Quality and of United Who's Who. She holds a Master's degree in Probability and Statistics from Auburn University and a Bachelor of Science degree in Mathematics and Computer Science from the University of North Alabama.

Dr. Boetticher's research interests include data mining (financial, bioinformatics, software engineering), machine learning optimization (genetic programs), and grid computing. He was a university finalist for the Minnie Stevens Piper award (teaching excellence) for the last four consecutive years. He has served on various conference and workshop program committees and is currently the co-chair for the 2006 Predictor Models in Software Engineering (PROMISE) Workshop.

Dr. Boetticher has been working in advanced machine learning applications in corporate, government and academic settings since 1986. He has served on the Executive Board Reuse Interoperability Group and the Executive board for the IEEE Reuse Standards committee.

The primary mission of the office is to engage in continuing research in all pertinent aspects of Baylor University and to provide complete testing services for Baylor students and public. Prior to this position, Dr. Bohannon served as the Director of the Office of Institutional Research at Appalachian State University and as an Associate Professor of Mathematical Sciences at Appalachian State University.

Dr. Bohannon has served in leadership positions in several professional organizations including AIR, SAIR, NCAIR, and TAIR. He has presented papers and workshops at the following professional associations for over 20 years—AIR, SAIR, NCAIR, TAIR, ASA, SUGI and others. His areas of specialization include Statistics, Data Warehousing and Data Mining, Institutional Research and Computer Application, SAS Training and Applications, Microcomputer Training Applications.

Dr. Bohannon earned a Ph.D in Statistics from Texas A&M University in 1976 and an M.A. in Mathematics from Wake Forest University in 1965. He also holds a B.S. in Mathematics with a Physics minor from McNeese State University.

Gary D. Boetticher is an Associate Professor in the Department of Computer Science and Software Engineering at the University of Houston—Clear Lake. He attained his Ph.D. in computer science from the West Virginia University.

Tom Bohannon, Ph.D. is the Assistant Vice President and Director at the Office of Institutional Research & Testing Services at Baylor University.

Randy Collica received a B.S. degree in electronic engineering from Northern Arizona University in 1982. He has 16 years experience in the semiconductor manufacturing industry working on yield and product and quality engineering. Since 1998 he has been with Compaq and Hewlett-Packard as a Sr. Business Analyst using Data Mining techniques for targeted marketing and customer analytics in the Customer Data & Knowledge Services department. He has developed customer scoring models and models to estimate corporate IT spending for use in tactical and strategic customer and prospect business intelligence. His current interests are in Clustering and Ensemble models, knowledge and data engineering, missing data and imputation, and text mining techniques for use in business intelligence. He has authored and co-authored 11 articles and is currently writing a book on CRM Segmentation and Clustering for business applications using SAS® Enterprise Miner™. Mr. Collica has been a member of the IEEE since 1979.

Dr. Goutam Chakraborty has a B.Tech (Honors) in Mechanical Engineering from Indian Institute of Technology, Kharagpur, a PGCGM from Indian Institute of Management, Calcutta, a M.S in statistics and a Ph.D. in marketing from University of Iowa. He has held managerial positions with a subsidiary of Union Carbide, USA and with a subsidiary of British American Tobacco, UK. He is currently an associate professor of marketing at Oklahoma State University where he has taught digital business strategy, web-business strategy, electronic commerce and interactive marketing, data mining and CRM applications, data base marketing, new product development, marketing research, and product and pricing management for the past fifteen years. He has presented numerous programs and workshops to executives, educators, and research professionals in U.S., Europe, Singapore, Hong Kong, Dubai, Abu Dhabi, and India. He has won many teaching awards including "Regents Distinguished Teaching Award" at OSU, "Outstanding Direct Marketing Educator Award" given by the Direct Marketing Educational Foundation, New York, Professor of the Year Award at CIMBA Italy, and Great Executive MBA Instructor award at the University of Iowa, Iowa City. His research has been published in many scholarly journals such as *Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, Industrial Marketing Management* etc. He has co-authored the book *Contemporary Database Marketing*. In addition, he serves(d) on the editorial review board of *Journal of Business Research* and *Journal of Academy of Marketing Science*. He has chaired the national conference for direct marketing educators for 2004 and 2005. He has also consulted extensively on issues related to developing digital business strategy, building and managing customer relationships, product development, and management and creation of e-business models with companies such as Aetna, Mercruiser, Thrifty Rent-A-car, Berendsen Fluid Power, Globe Life Insurance, Vanguard Realtors etc. For more information see his web page located at: **spears.okstate.edu/~goutamc**.

HungChing Chan is the Director of Clinical Analysis and Program Evaluation at Medica, a Minnesota-based health plan serving 1.3 million members. She has been at this position for six years. Her accountability includes all the data and analytical support for disease management programs, clinical quality improvement programs, and care management programs. Her department is also responsible for producing HEDIS (a standard health plan report card), physician profiling, program evaluation, and ROI analysis. HungChing received her bachelor's and master's degree in public health administration from National Taiwan University. She completed all course work and exams toward a doctoral degree. in Health Services Research, Policy and Administration from the University of Minnesota. Prior to joining Medica, she worked at Wellmark Blue Cross and Blue Shield of Iowa, Iowa State University, and Institute for Health Services Research at the University of Minnesota. Her research interest is in the quality of health care and measurement of performance.

Manoj Chari is the Analytical Solutions Manager in the Operations Research and Development department at SAS Institute in Cary, NC. He has a Ph.D. in Operations Research from the University of North Carolina at Chapel Hill. Before joining SAS in 2003, he was an Associate Professor of Mathematics at Louisiana State

University, Baton Rouge. He has over a decade of academic research and teaching experience in discrete mathematics and optimization. He has held visiting academic positions at the University of Waterloo, Canada, Technical University of Berlin and Indian Institute of Science. His professional interests are in optimization modeling, theory and algorithms, particularly in the area of large scale linear and combinatorial optimization.

Joseph Yechong Chen, Ph.D. is currently Vice President, Risk Modeling at Origen Financial. Origen is one of the biggest lending institutions in the manufactured home industry. Dr. Chen oversees the Risk Management Department at Origen. He has more than seven years of experience in risk management, quantitative modeling, and financial analysis in the financial industry and energy industry. He has been an invited speaker at Credit Risk Conferences by Incisive Media (previously Risk Waters Group). Previous positions held include Manager of Quantitative Analytics at Cinergy, Quantitative Analyst at Williams, Economist at Freddie Mac. Dr. Chen holds a Ph.D. in Management from the Business School of UCLA, and a M.A. in Physics from CUNY.

Cali M. Davis, Ph.D. is the Enrollment Research and Assessment Analyst in Enrollment Management at the University of Alabama (UA). Cali has utilized data mining and statistics to create tools for enrollment prediction, recruitment strategies, and at-risk student retention studies. She has presented professionally at the regional and national level. Prior to her current position, Cali taught statistics at the university and community college level. Cali received both her Ph.D. and Master's in Applied Statistics from the University of Alabama.

David Duling is the software development manager for SAS® Enterprise Miner™. He has worked at SAS since 1996 developing various components for EM including the process flow diagram, scoring functions, neural network, ensemble models, and link analysis. David has degrees in Physics and Statistics and previously worked for the National Institutes of Health developing and publishing methods of numerical analysis in simulations of magnetic resonance spectroscopy.

Abdu Elnagheeb is a Senior Vice President at Bank of America. He is the Manager of the Data Mining and Modeling Group in Analysis and Information Management (AIM) supporting Consumer Real Estate. Before that Abdu held similar positions within AIM supporting Credit Card acquisition and Consumer Debit Card and Deposits. His work includes Segmentation, Statistical Modeling, and development of economic logic to target and keep the most profitable customers for the Bank. Abdu also worked at American Express supporting both acquisition and portfolio management of Small Business credit card. Before that Dr. Elnagheeb worked as a Research Associate at the University of Georgia where his research focused mainly on Contingent Valuation Method, Water Resources, and Marketing. He has over 18 published papers in refereed journals, a book chapter, and several published abstracts. He attended and presented papers at a number of Regional, National, and International professional conferences.

Dr. Elnagheeb holds a Ph.D. and Masters degrees from University of Wisconsin—Madison with major in Economics and minor in Statistics/Quantitative Methods. He has extensive experience in the area of Statistical Modeling, Experimental Design, Data Mining, and their application in the Financial Sector, and over 15 years of SAS® programming experience.

Usama Fayyad is Yahoo!'s chief data officer and senior vice president of the Strategic Data Solutions group. Fayyad is responsible for Yahoo!'s overall data strategy, architecting Yahoo!'s data policies and systems, prioritizing data investments, and managing the company's data processing infrastructure and analysis. The analysis and management of data within Fayyad's group enables Yahoo! to ensure customer satisfaction and develop innovative products and services that are essential to consumers' lives.

Prior to joining Yahoo!, Fayyad co-founded and led the DMX Group, a data mining and data strategy consulting company. In early 2000, he co-founded and served as CEO of digiMine Inc. (now Revenue Science, Inc.), a data analysis and data mining company that built, operated, and hosted data warehouses and analytics for some of the world's largest enterprises in online publishing, retail, manufacturing, telecommunications, and financial services.

Fayyad's professional experience also includes five years spent leading the data mining and exploration group at Microsoft Research and building the data mining products for Microsoft's server division. From 1989 to 1996 Fayyad held a leadership role at NASA's Jet Propulsion Laboratory (JPL) where his work in the analysis and exploration of scientific databases gathered from observatories, remote-sensing platforms, and spacecraft garnered him the top research excellence

award that Caltech awards to JPL scientists, as well as a U.S. Government medal from NASA.

Fayyad earned his Ph.D. in engineering from the University of Michigan (1991), and also holds B.S.E.'s in both electrical and computer engineering (1984); an M.S.E. in computer science and engineering (1986); and an M.Sc. in mathematics (1989). He has published over 100 technical articles on the field of data mining, edited two influential books on the subject, and launched the primary scientific journal in the field while serving as editor-in-chief. Fayyad also launched and edited the primary newsletter in the technical community, and regularly delivers addresses and keynotes at government, industry, and academic conferences around the world.

George Fernandez is a Professor of Applied Statistics and the statistician for the Nevada Experimental Station and Cooperative Extension. He has more than 18 years experience in teaching introductory and advanced data analysis and statistical modeling courses using SAS software. He is also a professional SAS programmer and has over 25 years SAS programming experience. His major career accomplishments include winning best paper and poster presentation awards at regional and international SAS user conferences; presenting full-day workshops on applications of user-friendly statistical methods in data mining, being the keynote speaker at the 16th KSU Conference on Applied Statis-

tics, and organizing Western Users of SAS software (WUSS 7) conference in 1999. Many national and international SAS data miners are using his user-friendly SAS macro applications included in his book entitled *Data Mining Using SAS Applications* published by the CRC/Chapman Hall.

Dr. Michael Hardin is the Director of the Institute of Business Intelligence with the College of Business at the University of Alabama, as well as a professor of statistics at the University. He has authored or co-authored over 80 papers in various journals including the *Lancet, the Journal of the American Medical Association, the Journal of the American Medical Informatics Association, the American Journal of Epidemiology, the American Statistician, the Journal of Statistical Computation and Simulation*, and *Communications in Statistics*. He is the author or co-author of over 150 abstracts presented at national meetings and has given over 75 invited lectures or talks. He is the author of several book chapters dealing with database design and decision support systems.

Dr. Hardin often serves as a consultant to healthcare organizations in the areas of data mining, sampling, and program integrity. Additionally, he is an instructor and consultant for SAS in the areas of data mining and time series analysis. He is Adjunct Professor of Biostatistics and Adjunct Professor of Health Informatics at the University of Alabama at Birmingham.

He has served as Scholar in Residence in the Center for Information Management, Department of Information Systems and Operations Management, Loyola University, Chicago, and Visiting Professor in the Department of Management and Information Sciences and Statistics at Trinity College, Dublin, Ireland.

A member of numerous professional associations including the American Statistical Association, the Biometric Society, and the Institute of Mathematical Statistics, Hardin's specialty areas include data mining and knowledge discovery, data visualization, data warehousing, machine learning, statistical classification models, data management and collection methodologies, research design, informatics, the applications of statistical methodologies in the study of aging, and biostatistics.

Hardin earned a B.A. from the University of West Florida, an M.S. from Florida State University and an M.A. and Ph.D. from the University of Alabama.

William C. Hardy is a Senior Metrics Analyst in the Advanced Systems & Concepts Division of SAIC. He has more than thirty-five years experience as an operations analyst, divided evenly between focus on military communications, command and control and intelligence systems and commercial telephony and decision support systems. By virtue of publication of two books on the subject, a four year tenure as a regular columnist in the *QSDG Magazine*, published by the ITU Quality of Service Development Group, and 17 patents for innovations in test technology, he is an internationally recognized expert in measurement and evaluation of quality of voice telephony. He has also designed and developed, or directed development of, highly effective automated decision support systems for both military and commercial applications that have been fielded with a fraction of time and effort normally associated with such efforts. He holds a Ph.D. in Mathematics from the University of New Mexico.

Maria Herlihy has over 15 years' experience on the development and management of enterprise-wide database, analytical and marketing solutions. Her approach has centered around applying advanced analytics to large volumes of data for the purpose of solving specific complex business problems, surfacing latent opportunities and improving the ROI on marketing efforts. She has successfully managed database marketing and analytics services teams that provide ongoing consultation for the design and development of customer data warehouses. Her experience and expertise includes a wide variety of analytical and consulting functions from model development for direct marketing to inventory management, call center staffing and warehouse logistics.

Dr. Mark E. Johnson is Professor of Statistics at the University of Central Florida. He was formerly Department Chair and is currently on sabbatical leave at Northwestern University. Dr. Johnson is a Fellow of ASA, Chartered statistician of RSS and an elected member of the ISI. He has served on such editorial boards as *Technometrics, Journal of Quality Technology and Operations Research*. He has received research awards including the Jack Youden Prize, Shewell Award and the Thomas Saaty award. Dr. Johnson is an active consultant in the areas of hurricane modeling, data mining, sampling and has taught black belt statistics courses for IBM and Motorola. He has also testified in numerous administrative hearings as an expert in sampling. His consulting clients include BlueCross BlueShield of Florida, Lockheed Martin, the Agency for Health Care Administration, the Florida Commission on Hurricane Loss Projection Methodology and Kinetic Analysis Corporation (**hurricane.methaz.org**).

Mr. Paul B. (Brad) Jordan has been with BCBS of Florida for over 5 years. During that time he has served as the technical lead of Medical Informatics that analyses and forecasts claims experience that exceeds $3 Billion annually. In addition, he as administered SAS in a multi-platform architecture with over 500 users. In his current role, he is responsible for Marketing Analytics including the implementation of a consumer marketing database that provides for end-to-end campaign management, the creation and

validation of segmentation strategies for members, groups, physicians and agents, and for the development of customer lifetime value algorithms and applying them to the consumer marketing database to better understand cross-sell and value propositions for target customer segments. Brad has over fifteen years of experience with the SAS system. He has developed and facilitated instructor-based SAS training courses for nearly ten years for various clients. He also has practical experience using SAS in a variety of industries such as insurance, banking, finance, manufacturing, education and government. His management of computer programming staff offers the perspective of technical capability, user requirements, and organizational goals. He has worked in a number of industries where he has been responsible for design and implementation of SAS-based applications including reporting, forecasting, decision support systems, data marts and predictive modeling/data mining. In addition to his work experience, Brad has served as the head of several SAS user groups, is a Certified SAS professional and has a B.S. in Information Management from Jacksonville University where he is currently pursuing an M.B.A.

Talbot Katz is a New York City based data mining consultant who has over a baker's dozen years of experience with quantitative marketing applications in the financial services, IT, pharmaceutical, and telecommunications industries. He has presented his recent research into optimal methods for clustering and predictor variable transformation at the SRI Artificial Intelligence Seminar in 2005, and at the 2005 and 2006 SIAM Data Mining conferences. A SAS enthusiast, Talbot has addressed audiences at NESUG, NJSUG, and NYASUG meetings. He has an A.B. in Mathematics from the University of California, Berkeley, and a Ph.D. in Mathematics from The Rockefeller University.

Colonel William J. (Bill) Kelley returned to active duty in January 2002 to support Operation Noble Eagle. He is serving as both the senior military officer in the Office of the Inspector General of the Department of Defense, as well as Program Director for Data Mining Division. Before January 2002, he was a Program Manager with the Air Force Audit Agency's Special Projects and Financial Support Division, March Air Reserve Base, California.

Colonel Kelley has been the OIG DoD focal point for the data mining of DoD purchase card data and in planning the joint audit to research potentially fraudulent or improper transactions. Colonel Kelley has coordinated with numerous DoD criminal investigative and audit organizations, the General Accounting Office, and the President's Council on Integrity and Efficiency Inspection and Evaluation Committee to support the Secretary of Defense's efforts to identify whether fraudulent or inappropriate charge card program activities have, in fact, occurred.

As Program Manager for the Air Force Audit Agency, he was responsible for planning, developing, programming, and applying Air Force-wide audits relating to Air Force comptroller issues. Specific subjects include military and civilian pay benefits, program and budget, and other comptroller functions to include travel, cash management, financial certifications, accounts receivable, accounts payable, overseas banking operations, reimbursements, and use of special funds. In addition, he participated with the Fraud Working Group (which included OIG DoD, Defense Criminal Investigative Service, Defense Finance Accounting Service, Air Force Audit Agency, Air Force Office Special Investigations, Army Audit Agency, US Army Criminal Investigative Division, Naval Audit Service, and Naval Criminal Investigative Service) on fraud. He

was also a member of the Air Force task force to develop a fraud prevention tutorial and an Air Force financial management self-inspection checklist program.

Before joining the Army Audit Agency, in August 1974, Colonel Kelley was active duty Army. After his discharge in 1972, he continued to participate in the Army National Guard. Among his many military accomplishments, he commanded two direct support artillery battalions (one in Texas, the other in California) and the 40th Infantry Division Artillery. Further, he was the president of the National Guard Field Artillery Council for 2 years. Colonel Kelley began his civilian career as an auditor with Army Audit Agency at the El Paso Audit Office and then transferred to the Fort Hood Audit Office in 1979 as Managing Auditor. In June 1985, Colonel Kelley became an audit manager with Air Force Audit Agency.

He holds a B.B.A. in accounting from the University of Texas in El Paso, is a graduate of the Armed Forces Staff College, and is both a Certified Information Systems Auditor and a Certified Information Security Manager. Also, he is a Fed 100 award winner for 2005.

Dmitri Kuznetsov is Senior Brand Analyst at Media Planning Group. He develops advanced mathematical, econometrical, and statistical predictive models for MPG clients. Prior to this, Dmitri worked for Sigma Marketing Group as a Senior Statistical Consultant in expert team, special-

izing in marketing, retail pricing, risk, and financial data analysis, research, forecasting, and optimization of decisions. Before Sigma, Dmitri worked for KSS Group, a professional software company in the UK, as a Senior Scientific Developer of advanced retail pricing algorithms. Prior to KSS Dmitri was a Quantitative Business Analyst at Giant Eagle, where he developed in-house models on risks, marketing and price optimizations for revenue and profit management.

Dmitri received both his Ph.D. and Master's Degree in Mathematics and Physics from Moscow State University in Moscow and has worked as principal researcher on several initiatives with the NATO, International Science Foundation, Russian Foundation for Basic Research, and US Civilian Research Development Foundation. His academic experience also includes research at University of Pittsburgh in Pennsylvania and University of Waterloo in Canada.

Richard W. La Valley is a Senior Metrics Analyst in the Advanced Systems & Concepts Division of SAIC. He has more than thirty years experience as a statistician and operations analyst working in the areas of sample survey design, decision support, command and control, trouble tracking and monitoring systems in both commercial and government market segments. He has also directed and designed real-time development fraud detection tools for telephony and has developed numerous highly effective decision support systems which have been fielded in commercial telephony. He holds a Masters degree in Statistics from Pennsylvania State University.

of all individual under 65 sales at Blue Cross Blue Shield of Florida. In addition to her BCBSFL experience, Ms. Locke has 11 years of Direct Marketing experience in both the financial services and retail industries. In positions prior to BCBSFL, her roles have included:

• Developing new distribution and response channels in order to increase market share,

• Developing new marketing creative, resulting in a significant increase in response,

• Managing a Marketing Database, including maintenance, growth analysis, list selection, cluster analysis, modeling and segmentation.

Ms. Locke has a B.S., Business Administration from the University of Florida and is licensed as an insurance agent in the state of Florida.

Kim Larsen the director of the Data Mining and Statistical Analysis group at Charles Schwab & Co., in San Francisco. His focus is on client segmentation as well as predictive modeling to reduce client attrition and maximize campaign response rates. Prior to joining Schwab, Kim worked in home equity credit risk management.

Kim holds a B.S. in mathematics and economics and an M.S. in statistics.

Ms. T. Lynn Locke joined Blue Cross Blue Shield of Florida in 1999. She is currently leads all direct marketing activities consumer products at CBSFL. In this role, she is responsible for Promotional campaign strategies and execution, Campaign analysis and reporting, Sales analysis and forecasting as well as for Optimization of member growth strategies. Prior to her current role, Ms. Locke served as Project Manager and Business Lead for the implementation of a Web Sales capability that now accounts for 14%

Dr. Jing Luan is Chief Planning, Research & Knowledge Systems Officer at Cabrillo College on the beautiful coast of Monterey Bay next to Silicon Valley. He is responsible for reporting, assessment validations, strategic master planning, and accreditation planning. His additional interests include Knowledge Management, Data Mining and Data Warehousing with the emphasis on web based applications and access. His experience and expertise range from strategic planning, educational services, information management, and research to benchmarking. He chaired a data warehousing project of 109 community colleges in California, the world's largest higher education data warehouse.

He has held executive and leadership positions on a number of national and state committees and organizations. He was President of the RP Group (2001-2002), President of California Association of Institutional Research (CAIR, 2002-2003), Chair of Overseas Chinese Association of Institutional Research (OCAIR, 2000-2001), President of NCCCRP (National Community College Council for Research and Planning, 2004-2005), and Chair of Governor's School-to-Career Evaluation and Accountability Committee (1999–2002). He was appointed to serve as a member of the board for the iJournal, as data facilitator for the national Achieve the Dream initiative, and a member of Campaign for California Community Colleges.

Dr. Jing Luan is a well-published author on a variety of subjects in higher education and information technology in general. He co-authored the book, *Knowledge Management—Building A Competitive Advantage in Higher Education*, published by Jossey-Bass. He has conducted many professional development workshops, including student services student learning outcomes assessment. He has held many data mining workshops with high attendance and good reviews. He was voted Best Presenter at American Association of Institution Research conference in Seattle, Washington and was the Recipient of Best Paper award from CAIR.

He holds a Ph.D. in Higher Education Management & Policy Studies from Arizona State University, a Master's in Curriculum and Instruction from New Mexico State, a B.A. in English Literature from Shanxi University, PR China, and Certificate of Information Technology Management from University of California at Santa Cruz Extension (UCSC).

James (Jim) W. Mentele graduated from South Dakota State University with a B.S. in mathematics and a M.S. in physics. He worked for the Dow Corning Corporation from 1966–2001, retiring as a Senior Information Scientist. He has been working with the Central Michigan University Research Corporation as a Senior Research Fellow since then. He also works with Global Language Translation and Consulting (GLTaC).

Will Neafsey is the Consumer Segmentation and Analytics Manager for Ford Motor Company. During his last 14 years with Ford, he has worked in Market Research, New Business Creation and Incubation, Information Technology, Operations Research, Manufacturing, and at Ford Financial. He holds a Bachelors and Masters Degree in Operations Research from Cornell University, as well as an M.B.A. from Cornell's Johnson School.

Dr. Daniel C. Payne is the principal epidemiologist and project officer for the Centers for Disease Control and Prevention (CDC), Vaccine Analytic Unit (VAU). In collaboration with the Department of Defense and the Food and Drug Administration, the VAU conducts vaccine post-marketing surveillance and epidemiological analysis on vaccines administered in the US military using a large, linked health and vaccine database, the Defense Medical Surveillance System (DMSS). Completed studies describe the vaccine safety research agenda processes, the relationship between anthrax vaccine and optic neuritis, the risk of hospitalization associated with the military practice of multiple near-concurrent vaccinations, and a vaccine data quality assessment of the DMSS.

Before leading the VAU, Dr. Payne completed a CDC post-doctorate epidemiology fellowship developing pathology surveillance systems and a bioterrorism surveillance guidebook for medical examiners. Prior to joining the CDC, he was the State Director for the Louisiana Covering Kids Initiative, a Robert Wood Johnson Foundation grant to improve child health care access, and served Louisiana and Maryland as a surveillance epidemiologist.

Dr. Payne received his Ph.D. from Tulane University, Graduate School of Medical Sciences, a M.S.P.H. degree from the University of Alabama, Birmingham (UAB), School of Public Health, and a B.A. from Emory University in Atlanta, Georgia.

Greg Potts is the Data Mining Team Leader for the Arkansas Foundation

for Medical Care. As Data Mining Team Leader, Mr. Potts applies tools such as SAS® Enterprise Miner™ to Medicaid claims data in an effort to identify key drivers of health care costs in the Arkansas Medicaid population. Mr. Potts has also served as the state Coordinator for the Behavioral Risk Factor Surveillance System (BRFSS) survey with the Arkansas Department of Health, and has managed several survey research projects from inception to completion with other companies.

Uzma Raja is an Assistant Professor of MIS at the Manderson Graduate School of Business, the University of Alabama. Her area of research is software system development and maintenance. Her research on Open Source Software projects using Data Mining and Text Mining has been presented and published at various conferences and professional meetings, including SUGI 31, INFORMS and ICSE.

Dr Raja earned her Ph.D. in Information Systems from Texas A&M University. She also holds a B.S in Electrical Engineering from the University of Engineering and Technology, Lahore, Pakistan, an M.S. in MIS from Texas A&M University. She has over seven years of experience as a systems engineer with a global consulting company.

Timothy Rey graduated in 1979 with an M.S. in Forestry Biometrics from Michigan State University. He joined Dow in the summer of 1979 in the Math Applications group in Central Research as a Research Statistician. While in the Math Applications Group, Tim conducted projects, consulted, taught and developed linear statistical methods at most of Dow's North American sites. From there he went on to supervise the Math Applications group in 1985 and then on to manage the Research and Technical Applications group in Central Research in 1987. In 1988 he took on the technical management role for the SimuSolv nonlinear simulation, optimization and estimation commercial software development effort in Dow's Central Research organization. Here he conducted joint research with universities to develop nonlinear statistical methods for experimental design and maximum likelihood estimation. At that time he was also responsible for Dow's Quality program in Central Research. In 1996 Tim moved to Dow's Marketing and Sales Expertise Center where he lead the development of Dow's Customer Loyalty program and was also the lead Marketing Research process owner. In 2005 Tim became the manager of Dow's Data Mining and Modeling group situated in Dow's Six Sigma Expertise Center. Tim as written over 100 internal papers as well as published 15 papers externally. He has delivered numerous talks at various quantitative methods forums.

Olivia Parr-Rud is an internationally known Business Intelligence expert, speaker and author of *Data Mining Cookbook* (Wiley 2001). She has over 15 years experience in marketing analytics and business intelligence for a variety of industries. Her passion for creating successful solutions for her clients is guiding her research into the organizational competencies and issues that are critical to the success of business intelligence projects.

Through OLIVIAGroup, she offers a combination of services designed to empower companies to maximize their marketing potential through strategic consulting, predictive modelling and data mining. To support her research, she is pursuing a Ph.D. in Human and Organizational Development through Fielding Graduate University. And she is a founding partner of IntegrityIQ, a company dedicated to helping organizations build congruence around their vision and mission through a unique tool, the Organizational Integrity Survey. She chaired the Data Mining Conference in 1999.

Olivia is an instructor for SAS Institute in the Business Knowledge Series. She teaches a 2-day course based on her book, *Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management* (Wiley 2001), now available in 4 languages.

Olivia also serves as faculty member for Data University, an online university that features experienced database marketing professionals who provide one-on-one training worldwide via the Internet. She has just completed a home study course in predictive modeling with SAS.

Olivia has held senior management positions at Fleet Bank, National Liberty Insurance and Providian Bancorp. She has a B.A. in Mathematics from Gettysburg College and an M.S. in Decision and Information Systems, with a concentration in Statistics, from Arizona State University.

Günter Schmölz, 36, was born in Vienna, Austria. He studied informatics, economics and insurance mathematics at the Technical University of Vienna and at the University of Economics in Vienna. Günter Schmölz has been working in Insurance business for more than 10 years. He started at Allianz Insurance in South-Africa. During his time with Allianz Insurance Austria he became part of the Austrian actuarial society. When he worked for AXA Insurance Austria (2002-2003) he was responsible for reinsurance and the non-life actuarial department. In 2004 Günter Schmölz changed to Uniqa Insurance Austria and became head of the department "Customer Intelligence". During his time in insurance business Günter Schmölz mainly focused on Tariff Calculation, Analytical CRM and Predictive Modeling. Günter Schmölz lives in Vienna, is married and is father of two children.

Mr. Schwartz has more than 25 years of experience as a hands-on enter-

prise software technology expert and team builder. He has served four times as CTO for small to medium sized startups where he has led organizations to successful product delivery, revenue, market differentiation and funding. Schwartz holds a Masters Degree from Brooklyn College and a B.A. from the University of Delaware.

While CTO/SVP Engineering at Passport Corporation, 2001-2005, Schwartz conceived and developed their wireless architecture and protocol. Later Schwartz served as principle architect and key developer of Passport's flagship wireless time management product, inTime for Kronos.

Prior to Passport Corporation, Schwartz was VP Software Development and later CTO of Skila Corporation, 1998-2001, a healthcare marketing intelligence ASP. While at Skila, Schwartz conceived of an innovative technology (later named "Intelligration Plug-in") that allowed any Pharmaceutical or Medical Device BI researcher to click a button in any Web browser to highlight and drill-down on the details of any healthcare related topic. This technology was submitted to the USPO for a software technology patent.

Prior to Skila, Schwartz was CTO/VP Engineering of InSync Corporation, where he was responsible for the architecture and innovative design concepts that allowed all of the company's enterprise applications to automatically run on multiple graphical windowing platforms without rewriting.

Prior to InSync Corporation, Schwartz, worked as the CASE Technology Product Manager, for TeleSoft Corporation, a compiler company for embedded real-time Ada. In this role, Schwartz pioneered the ASIS standard, providing the first abstract semantic interface between the Ada programming language and multiple CASE platforms.

Prior to TeleSoft, Schwartz was CTO/VP Engineering for Fortex Corporation, 1987-1988, where he was responsible for redeveloping their portable, database independent, run-time architecture, extending their 4GL language, directing the development of their software automation tools, and porting their product to multiple operating systems and graphical windowing systems.

William (Bill) Smith is Executive Director of the American Statistical Association and Professor Emeritus of Statistics at Texas A&M University, where he was chairman of the Department for a period of nine years and served in A&M's College of Science Dean's Office for an additional nine years. Prior to coming to the ASA, Smith was a program director in statistics at the National Science Foundation. His research interests are in the area of multivariate analysis, especially with regard to applications to industrial, educational and legal processes. He has co-authored two books and over fifty articles in statistics.

He was editor-in-chief for the *Communications in Statistics* for ten years. Smith is a Fellow of the American Statistical Association, received a distinguished teaching award (university-level) at Texas A&M, was name a Distinguished Alumnus of the College of Science at Texas A&M University and he received the H. O. Hartley and Don Owen awards.

Judy Spomer has been a Member of Technical Staff at Sandia National Laboratories for 3 years. Sandia National Laboratories develops and applies advanced technology to keep our nation secure and capable of meeting the toughest challenges. At Sandia, Judy has developed models that have been used in forecasting safety incidents, development of ergonomics evaluations, and in the area of security. Prior to employment at Sandia, Judy worked as a Risk Modeler for American General Finance, developing customer credit scoring models, and behavioral models aimed at reduction of customer delinquency. Judy has a B.S. in Computer Science from Indiana University of Pennsylvania, and is pursuing an M.S. in Data Mining at Central Connecticut State University where she is a Graduate Teaching Assistant in the data mining program.


Gerhard Svolba, Ph.D., is programming in SAS since 1991. He worked at the University of Vienna, Department for Medical Statistics where he performed a lot of biometrical analyses and build numerous analysis tables for clinical trials. In his Ph.D. he focused on Statistical Quality Control in Clinical Trials, where he showed the application and data preparation for the control of quality characteristics in clinical trials.

Since 1999 he is working as a Consultant for SAS-Austria, where he is involved in numerous analytic projects in the CRM, Basel II and demand forecasting area across various industries. He has the product manager responsibility for the analytic SAS products and solutions in Austria and has put his hands on a number of analytic datamarts creations as well as creating concepts for analytic projects and data preparation.

Gerhard Svolba lives in Vienna, Austria. He is married, father of three sons. Beside working for and with SAS he likes to spend time with his family and to be out in the nature, especially sailing on a lake.
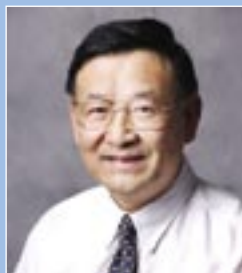
Thomas T.H. Wan, Ph.D., is Professor of Public Affairs, Health Services Administration, and Nursing at the College of Health and Public Affairs, UCF. He directs an interdisciplinary doctoral program in public affairs and serves as an associate dean for research at the College of Health and Public Affairs. Before joining UCF, he was Professor of the Department of Health Administration, Medical College of Virginia, VCU. He held the Arthur Graham Glasgow Chair (1991-1999) at VCU. Professor Wan received a Bachelor of Arts in Sociology from Tunghai University, Taiwan; a Master of Arts and a doctorate in Sociology/Demography from the University of Georgia, and a Master of Health Sciences from the Johns Hopkins University School of Public Health. He is a member of the Health Services Organization and Delivery Study Section, NIH. His research interests are centered in managerial epidemiology, health services evaluation, health informatics, and clinical outcome studies. His published work includes 8 books; *Analysis and Evaluation of Health Care Systems: An Integrated Managerial Decision Making Approach; Evidence-Based Health Care Management: Multivariate Modeling Approaches; Monitoring the Quality of Health Care: Issues and Scientific Approaches*, etc. and 100+ scientific articles. His project, entitled "Nurse Staffing & Nursing Home Quality," is funded by NINR/NIH (2003-2007). He is a guest editor on *Health and Public Affairs Informatics Research for the International Journal of Public Policy*, where he serves on the editorial board.

Morgan C. Wang received his Ph.D. in 1991. Since then he has been with UCF, where he is currently the Director of Institute of Statistics and Data Mining and Professor of the Department of Statistics and Actuarial Science. He has published a book *(Integrating Results through Meta-Analytic Review Using SAS® Software*, SAS Institute, 1999), and over 50 papers in referee journals and conferences on topics including interval analysis, meta-analysis, and data mining. He was the first prize-winner in Data Mining Competition of the 11th SIGMOD KDD conference in 2004 and the first prize winner in Data Visualization Contest of SUGI 25 conference in 2000, and was given invited talks for more than

fifty times for American Statistical Association, Society for Industrial and Applied Mathematics, Iranian Statistics Association, and many universities around the world. Dr. Wang is an active consultant in data mining and business intelligence. His consulting clients include Blue Cross Blue Shield of Florida, AAA Auto Club South, Sodexho Campus Services, Florida Department of Health, University of Central Florida and United Space Alliance.

Cary White is the Data Warehouse Development Team Project Manager for the University of North Carolina at Chapel Hill. He has over 20 years of experience as an IT developer and project manager. For the last four years, Cary has worked on the development of an enterprise-wide data warehouse for UNC-CH. Prior to that, he designed, developed, implemented and supported HR, Payroll and Accounting applications. In a previous career, he was an instructor at a small college. Cary holds an MA in Sociology from Duke University and a B.A. from Davidson College.

Tom Winn is a mathematician by education, with more than 31 years of experience working in state government. He has been a SAS user for more than 23 years, and is President-Emeritus of the South-Central SAS Users' Group. Tom currently is a Senior Systems Analyst at the Texas State Auditor's Office, in Austin, Texas, where his work involves data analysis and the development of intranet applications.

Terry Woodfield is a Statistical Services Specialist in the Education Division of SAS and served as co-chair for M2003, SAS' 6th annual data mining conference. Dr. Woodfield has more than 29 years of SAS programming experience and has provided training and mentoring services in the areas of statistical forecasting, predictive modeling, and data mining. At SAS, Dr. Woodfield has developed courses in statistical forecasting, Web mining, and text mining. He is also active in the statistics profession, presenting papers at numerous statistical conferences and professional meetings, and he has served on steering committees in data mining and forecasting. He has helped develop forecasting and predictive modeling solutions for insurance, energy, and retail companies and been an expert witness in utility ratemaking hearings. Before joining SAS, Dr. Woodfield was Chief Statistician at HNC Software and other prior experience includes statistical software development in SAS/ETS Research and Development and university teaching and research.

Jeff Zeanah is the President of Z Solutions, Inc. a firm focused on the support of organizations through predictive analytics and exploratory data mining. His primary interests and research concern the problems organizations face to improve their business decisions through data analysis, including predictive analytics and the selling of the results. Jeff has consulted with industry leaders in manufacturing, retail, public health, science, finance, nutrition and utilities. He is an instructor for SAS Institute Inc. A frequent guest lecturer at universities on the topic of applying analytics to business, he serves on the board of the Institute for Business Intelligence at The University of Alabama.

As a recognized expert on neural networks and a broad range of exploratory data mining tools Jeff has authored papers on neural networks, exploratory data mining, and the implementation of those techniques in organizations. He is the developer of exploratory approaches and techniques that have been used worldwide by Fortune 500 companies, independent researchers, government agencies, and over 30 universities worldwide. His approaches have been applied in areas as diverse as improving manufacturing processes, analyzing market research, tasting wines, searching for oil, controlling river flow, sizing electric transformers, and classifying stars.

M2006 Exhibit Hall

Visit the M2006 Exhibit Hall and learn how our exhibitors' products and services can keep you on the cutting-edge in the field of data mining. The exhibit hall will be open throughout the duration of the conference. Below is a list of scheduled activities located in the exhibit hall.

Monday, October 23

7:30 a.m. – 8:45 a.m.	Breakfast
10:00 a.m. – 10:30 a.m.	Morning Break, Refreshments served in Exhibit Hall *(sponsored by Bank of America)*
11:30 a.m. – 12:30 p.m.	Lunch, Exhibit Hall open
3:45 p.m. – 4:15 p.m.	Afternoon Break, Refreshments served in Exhibit Hall
5:30 p.m. – 7:00 p.m.	Networking Reception in Exhibit Hall *(sponsored by SAS)*

Tuesday, October 24

7:30 a.m. – 8:45 a.m.	Breakfast
10:00 a.m. – 10:30 a.m.	Morning Break, Refreshments served in Exhibit Hall *(sponsored by CapitalOne)*
11:30 a.m. – 12:30 p.m.	Lunch, Exhibit Hall open
3:45 p.m. – 4:15 p.m.	Afternoon Break, Refreshments served in Exhibit Hall

Vendor Trivia

Participate in Vendor Trivia for your chance to win! In your conference bag, you'll find a game card with a list of questions. Simply visit each booth in the exhibit hall to find the answers to the trivia questions. Once your card is complete, drop it off at the registration desk to be entered into a drawing to win a variety of prizes.

Cards must be completed and turned in by 12:00 p.m. Tuesday, October 24 at the Registration Desk. Winners will be announced during the afternoon keynote address on Tuesday.

SAS Publishing

Visit the Bookstore on the exhibit hall and check out the latest offerings from SAS Publishing. SAS Publishing provides documentation to help meet the needs of new and experienced users of SAS software through an array of delivery methods, including e-books, CD-ROM, hardcopy books, and Web-based training.

Book Signing

Olivia Parr-Rud will be signing her book, *Data Mining Cookbook* during the Monday Evening Reception. Copies will be available for sale at the Bookstore.

Exhibitors and Sponsors

[x+1]
Contact: Jason Shulman
(212) 741-4222
jshulman@xplusone.com
www.xplusone.com

New York-based [x+1] helps marketers simplify their online marketing and achieve greater return on their overall marketing investment. We are reinventing the field of conversion optimization by combining insightful customer service from internet marketing experts with proprietary technology. Leading companies in financial services, telecommunications, on-line services, and automotive industries have significantly increased message accuracy, customer response and ROI with [x+1].

CSC
Contact: William Lunz
(630) 472-1469
wlunz@csc.com
www.csc.com

Computer Sciences Corporation helps clients achieve strategic goals and profit from the use of information technology. With the broadest range of capabilities, CSC offers clients the solutions they need to manage complexity, focus on core businesses, collaborate with partners and clients, and improve operations. CSC makes a special point of understanding its clients and provides experts with real-world experience to work with them. CSC is vendor-independent, delivering solutions that best meet each client's unique requirements. For more than 45 years, clients in industries and governments worldwide have trusted CSC with their business process and information systems outsourcing, systems integration and consulting needs.

Inductis
Contact: Brian Paris
(212) 284-3300
bparis@inductis.com
www.inductis.com

Inductis, an EXL Service company, is a global professional services firm that helps large companies leverage the information age to make better decisions through deep analytics. Inductis has a strong track record of helping leading companies strengthen their businesses with its unique mix of world-class mission critical business consulting and state-of-the-science analytical capabilities and infrastructure. We specialize in identifying high-impact problems and opportunities where we can generate results quickly, creating immediate and lasting value. We focus on value creation through our two practice areas: Management Consulting and Analytics Services.

Capital One
Contact: Craig Dye
(804) 347-8974
craig.dye@capitalone.com
www.capitalone.com

Capital One is a Fortune 200 company providing access to credit to tens of millions worldwide. We also provide auto, home and business loans, plus banking and other services. Statistical expertise is central to our success. Among thousands of scientifically trained professionals, over 100 statisticians ensure our governance and quality of statistics.

Hewlett-Packard
Contact: Janice Shineman
(978) 579-9708
janice.shineman@hp.com
www.hp.com

HP is a technology solutions provider to consumers, businesses, and institutions globally. The company's offerings span IT infrastructure, global services, business and home computing, and imaging and printing. For the four fiscal quarters ended July 31, 2006, HP revenue totaled $90.0 billion. More information about HP (NYSE, Nasdaq: HPQ) is available at www.hp.com.

Quantec, LLC
Contact: Robin Way
(503) 228-2992
robin.way@quantecllc.com
www.quantecllc.com

Quantec provides superior analytic and strategic consulting services. Our clients rely on us to provide comprehensive, meaningful solutions to complex problems. Quantec's team of economists, statisticians and engineers brings experience in large-scale project management, state-of-the-art modeling tools, and innovative research methods to every project. Our current experience includes work in the financial services, energy and manufacturing industries and the public sector. Quantec is a SAS Gold Alliance member and a leading provider of analytics-oriented professional services for SAS clients in North America and abroad.

Bank of America
Contact: Allison Reynolds
(704) 683-6951
allison.reynolds@bankofamerica.com
www.bankofamerica.com

Our company has embraced higher standards throughout our almost 200-year history, and all our work in recent years has been about pursuing higher standards—from our Six Sigma and process excellence work, to The Bank of America Spirit, to our Leadership Model, Diversity and Inclusion and our Code of Ethics and our many other customer and associate satisfaction initiatives. Higher Standards is simply one way we express what we already know about ourselves and our company.

Academic Pavilion

Oklahoma State University

Contact: Dr. Goutam Chakraborty
(405) 744-7644
Goutam.Chakraborty@okstate.edu
www.okstate.edu

The Department of Marketing at the Spears School of Business in Oklahoma State University is home for the SAS/OSU data mining certificate program. The curriculum for the is program has been carefully deigned to prepare Masters students with strong analytical and database management skills to work as a business analysts..

Center for Marketing Studies (MMR Program)

Contact: Richard Fox
(706) 542-3761
rfox@terry.uga.edu
www.terry.uga.edu/mmr

The Coca-Cola Center for Marketing Studies, established in 1986 at the University of Georgia, Terry College of Business, seeks to advance professional marketing education and to identify and develop new programs and methods which will better serve the marketing community's needs for education and information. The Master of Marketing Research Program, a nationally recognized graduate program within the Terry College of Business, is administered by the center. The center also works to maintain and strengthen relations with the business community.

Central Michigan University Research Corporation

Contact: Timothy Pletcher
(989) 774-2424
tim.pletcher@cmich.edu
www.cmurc.com

CMU-RC offers industry a source for innovation and collaboration with academia to resolve difficult business challenges. Our methodology coordinates access to faculty, experts, and students to undertake timely, low risk proof-of-concepts to learn about successful application of advanced analytics. CMU-RC leverages leading edge BI technology from SAS, IBM, and ESRI.

Institute for Business Intelligence and Applied Statistics Program
The University of Alabama

Contact: J. Michael Hardin
(205) 348-8901
mhardin@cba.ua.edu
www.cba.ua.edu/issms/

The Institute for Business Intelligence is an applied research center which serves as a key focal point for University/Industry partnerships in Applied Research, Business Intelligence and Data Mining. The Applied Statistics Program offers M.S. and Ph.D. degrees in Applied Statistics with specialized courses in data mining.

University of Central Florida Data Mining Lab

Contact: Morgan Wang
(407) 823-2818
cwang@mail.ucf.edu
www.cas.ucf.edu/statistics/

The award-winning Data Mining Program at the University of Central Florida offers an established educational environment complemented with ongoing industrial collaborations. Every graduate of the program has served an internship with an eminent Florida business partner in the aerospace, entertainment, insurance, hospitality or automobile service industry. Moreover, faculty have established consulting relationships with industrial clients inspiring relevant research directions, student employment opportunities and enhanced curriculum case studies. The Data Mining Program was officially established in 2000 at UCF, and the core faculty have had extensive experience in data mining and statistical consulting. UCF faculty and students won two 2004 KKD Cup Prizes.

University of Louisville

Contact: Adel S. Elmaghraby
(502) 852-0470
adel@louisville.edu
www.cs.louisville.edu
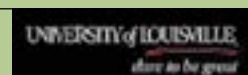
Faculty are leaders in Data Mining, Data Warehousing, and Data Analysis. Expertise and publications includes handling of text, web, numerical, and image data bases. Faculty funding comes from NSF, NASA, NIH, DHS, and DoD. The University of Louisville is designated Center of Educational Excellence in Information Assurance by the NSA.

SAS Pavilion

In the SAS Pavilion of the exhibit hall, you will find SAS Consulting, SAS Data Mining Offerings, and SAS Education. Also, be sure to visit the SAS demo theater to see the latest features and functionality available in SAS products.

SAS® Demo Theater

Visit the SAS Demo Theater and see first-hand how easy it is to build models in SAS Enterprise Miner. Both Beginners and Intermediate/Advanced users will benefit from these sessions. Visit our booth for schedules and details.

Demo Theater Schedule

MONDAY, OCTOBER 23
10:00 a.m. – 10:30 a.m.	Beginner Users
6:00 p.m. – 6:30 p.m.	Intermediate/Advanced Users

TUESDAY, OCTOBER 24
10:00 a.m. – 10:30 a.m.	Beginner Users
3:45 p.m. – 4:15 p.m.	Intermediate/Advanced Users

SAS® Data Mining Offerings
SAS® Consulting

Products: mary.crissey@sas.com
Consulting: zul.sidi@sas.com
www.sas.com/technologies/analytics/datamining
www.sas.com/consult/usconsult/analyticalconsulting.html

Product Demos

Visit the SAS Data Mining Software and Consulting booth to experience the most powerful, complete data mining solution on the market. Be the first to see what's new including the debut of our unparalleled model deployment solution. Check out the brand new Text Miner version 3.1 as it unlocks the hidden value of textual information. Come and experience how data mining can help you in your particular situation. Live demos will be given on demand for SAS® Text Miner, SAS® Enterprise Miner™, and SAS® Model Manager.

Analytical Consulting

Bring your questions and engage in a free hour of expert-to-expert analytical consulting. You will learn how the SAS Analytical Consultants can help you achieve your business objectives and leverage the expertise of SAS.

SAS Education

training@sas.com
support.sas.com/training/us
Learn to use SAS software from the people who know it best. SAS Education provides a full curriculum of SAS and Data Mining courses focusing on a broad range of topics including Business Knowledge Series classes—courses developed and presented by industry leaders.

Courses are delivered in our regional training centers, on-site at your location or via a Live Web class in real-time over the Internet.

M2006 RECEPTION
HOSTED BY SAS

Come mix and mingle with M2006 attendees in the Exhibit Hall, Monday evening, 5:30 – 7:00 p.m.

Enjoy food, drinks and live entertainment compliments of SAS.

All of the exhibit booths will be open, including the SAS Data Mining Offerings and Consulting booth where you can see live on-demand demonstrations of the most powerful, complete data mining solution on the market.

M2006 Poster Session

Located in the exhibit hall, visit the poster session to learn how these innovators in the data mining community are applying real methodologies, techniques and new ideas to their real-world work and research.

Data Mining the Price of Used Cars
Robert "Edward" Egros, Southern Methodist University

Property Prediction Model for a Refinery Process
S. Suganya

Generic Data Mining
R. Pradeep Kumar, University of Mysore

Affinity Modeling in a South African Bank
S. M. Millard, University of Pretoria

Exploring Market Basket Analysis Results
Patricia Cerrito, University of Louisville

A Risk Profile in the Swedish Social Insurance Agency
Cecilia Ekedahl, Forsakringskassan

Analyzing Buying Patterns at Retail Stores
Kuriakose Athappilly, Western Michigan University

Text Mining to Analyze Hospital Cost
Chakib Battioui, University of Louisville

Practical Implications of ROI Analysis
Alejandro Jaramillo, DataMeans

Data Mining to Investigate Prescription Practices
Mussie Tesfamicael, University of Louisville

A New Similarity Measure of Session Clustering
Maryam Ramezani, De Paul University

Data Mining Physician Decisions for MRSA
Hamed Zahedi, University of Louisville

Data Mining University Work Expectation
Guoxin Tang, University of Louisville

Neural Networks to Prevent Insolvency
Carlos Andre Reis Pinheiro,
Brasil Telecom/Coppe-UFRJ

student poster contest winner

Search for Multivariate Structure for Ecological Data Using Partition Modeling Approaches
Feng (Faith) Gao, Virginia Tech

Pattern Detection in Health Care Marketing Segmentation
Jin Su, University of Central Florida

Resource Optimization at Central Michigan University
Monica Mull, Central Michigan University

Web Crawling Using SAS® Enterprise Miner™
Sandra W. Cameron, University of Louisville

Improving Graduate School Student Quality
Anthony Vaiciulis, University of Central Florida

Predictive Modeling and Cluster Analysis Using Data Mining for an Insurance Company
Aswin Bhuvanendran, Oklahoma State University

Exploring the Relationship Between Classifier Performance and Dataset Complexity
Ana Landeros, The University of Alabama

An Application of Data Mining Formula
Hanmo Oh, Oklahoma State University

Spatial Allocation Simulator for Branch Network Planning
Chingyun Lin, U.S. Bank

Using Data Mining Approach to Assess the Financial Health of Higher Education Institutions
Ariel Huang, University of Houston

Predicting Rare Alarms on Mobile Telecommunications Networks
Luis E. Rocha Mier, SAS Mexico

Using Data Mining Tools to Predict a Rare Consumer Purchase Behavior: A Case Study
Shawn Rouen, Hallmark Cards, Inc.

Map of Rio Hotel Conference Center

Lunch
General Sessions/Keynotes
Exhibit Hall
Breakout Sessions
Registration

Service Corridor

Amazon
A   G

T
S
R
Q
P
O
N
M
K   L

B C D E F H I J

Public Corridor

Coco
A   B
Jaguar
A   B

Palma
A B C D
E F G H

Tropical
A B C D
E
F G H

Service Corridor

Expo Loading Dock

Expo Loading Dock

Public Corridor

Tango
A B C
Conga
A B C
Lambda
A B C
Choro

Dress'g Room
Dress'g Room
Load In

Men
Women

Lift

8   9   11
10

Elec Room
Dress'g Room

Stage

5   7
4   6
1   2   3

Brasilia

Load In

Public Corridor

Pavilion Kitchen & Heart of the House

5   6 Lift   7

Rio Pavilion

1   2
4
3

Load In
Bar   Part Store   Part Store

Public Corridor

7   8
5   6
Miranda
3   4
1   2

Women Men
Corridor
Elec Room
Ref
Kitchen
Office
Part Store
Part Store
Ref   Linen Store

Public Corridor

Flamengo Boardroom
Janeiro Boardroom

Registration
Rotunda
Registration

Bar
Coat Room
Storage
Men Women

Miranda Patio

Banquet Loading Dock

Conference Sales & Services Office

To Hotel

Pavilion Patio & Walkway

Passenger Drop-off

500,000 square feet of additional outdoor space

40

Conference Information

Conference Registration and Information Desk
Check-in and pick-up for your conference materials will be available from 11:00 a.m. – 5:00 p.m., Sunday, October 22 and 7:30 a.m. – 5:30 p.m. on Monday and Tuesday, October 23-24. The Registration and Information desk will be open both Monday and Tuesday during conference hours.

Breaks
Refreshment breaks are scheduled at 10:00 a.m. and 3:45 p.m. on both Monday and Tuesday. Refreshments can be found in the Exhibit Hall, Pavilion 5, 6, & 7.

Guests
Guests are not permitted in conference sessions. You must wear you conference badge at all times to be admitted into sessions.

Lunch
Lunch will be provided both Monday and Tuesday and is included in your conference fee. Lunch will be held in Amazon A.

Reception
The annual conference reception will be held Monday evening and is also included in your registration fee. Join your friends and colleagues in the Exhibit Hall on Monday evening, immediately following the last session and enjoy food, drinks and live entertainment, sponsored by SAS.

Smoking Policy
Smoking is prohibited in the Rio Conference Center. Smoking areas are conveniently located outside on the Miranda Patio.

Email/Internet
Attendees may use the Internet stations located in the Exhibit Hall and near the M2006 Registration desk to check email and access the Internet. Two of the stations in the Exhibit Hall are also connected to a printer for your use.

Security
Recording devices including video and audio are not permitted in the sessions.

Temperature Control
Room temperatures do fluctuate. Please bring a light jacket or sweater to ensure your comfort level.

Las Vegas Transportation

Free Shuttle to Harrah's from Rio Hotel
The Rio Hotel provides a free shuttle to Harrah's hotel, located on Las Vegas Blvd. (the strip). Pick-up/drop-off at the Rio is located at the entrance close to Carnival World Buffet. Pick-up/drop-off at Harrah's is located near the valet entrance at the bus departure section.

Shuttle service is available 10:00 a.m. – 4:00 a.m., Thursday—Sunday and 10:00 a.m. – 1:00 a.m., Monday—Wednesday. Pick-up times are approximately every 15 minutes. Sorry, no luggage is available on the shuttle.

Shuttle/Limousine Service

Bell Trans Shuttle Bus
702-739-7990
Operates from 7:00 a.m. – 2:00 a.m.

C.L.S.
702-740-4050
Shuttle and limo service
Operates 24 hours

Grayline/Coach USA/Express Shuttle
702-739-5700
Shuttle service
Operates from 7:00 a.m. – 1:00 a.m.

ODS
702-876-2222
Shuttle service
Sedan limo
Stretch limo
Operates 24 hours

Showtime
702-261-6101
Shuttle service
Operates from 4:00 a.m. – 12:30 a.m.

Taxi Service
Taxi cabs are continuously available for pick-up in front of every hotel entrance in Las Vegas.

M2006 Post Conference Training

If you are not signed up for a course, but are interested, seats may still be available. Please inquire at the Registration and Information Desk during the conference.

Registration for post-conference training opens at 7:30 a.m. on Wednesday, October 25. All classes start at 8:30 a.m. and end at 5:00 p.m.

Your instructor will inform you of lunch times daily. Lunch coupons must be redeemed by 1:30 p.m. Lunch is located in Amazon G.

Advanced Predictive Modeling Using SAS® Enterprise Miner™ 5.2

(This is a 2-day course offered Wednesday-Thursday, October 25-26.)
This course is designed for predictive modelers and data analysts who want to optimize the performance of predictive models beyond the basics.

Location: Amazon O-P

Applied Analytics Using SAS® Enterprise Miner™ 5.2

(This is a 3-day course offered Wednesday-Friday, October 25-27)
Designed for data analysts and qualitative experts as well as those with less of a technical background who want a general understanding of data mining, this course provides an introduction to data mining and an extensive hands-on experience with Enterprise Miner software.

Location: Amazon S

Data Mining Techniques: Theory and Practice

(This is a 3-day course offered Wednesday-Friday, October 25-27.)
This course is based on the newly revised and expanded book, Data Mining Techniques for Marketing, Sales and Customer Relationship Management. This course is for business analysts and their managers, statisticians, and anyone who has a professional interest in data mining.

Location: Amazon M-N

Decision Tree Modeling

(This is a 2-day course offered Wednesday-Thursday, October 25-26.)
This course is designed for predictive modelers and data analysts who want to build decision trees using SAS Enterprise Miner software

Location: Amazon Q

Design of Experiments for Direct Marketing

(This is a 3-day course offered Wednesday-Friday, October 25-27.)
This course is designed for market researchers who want to design experiments with more than one factor.

Location: Amazon T

Introduction to Survival Analysis Using Empirical Hazards

(This is a 2-day course offered Wednesday-Thursday, October 25-26.)
This course is designed for technically savvy business managers, SAS programmers and anyone who is interested in learning about applying time-to-event analysis to business problems.

Location: Amazon R

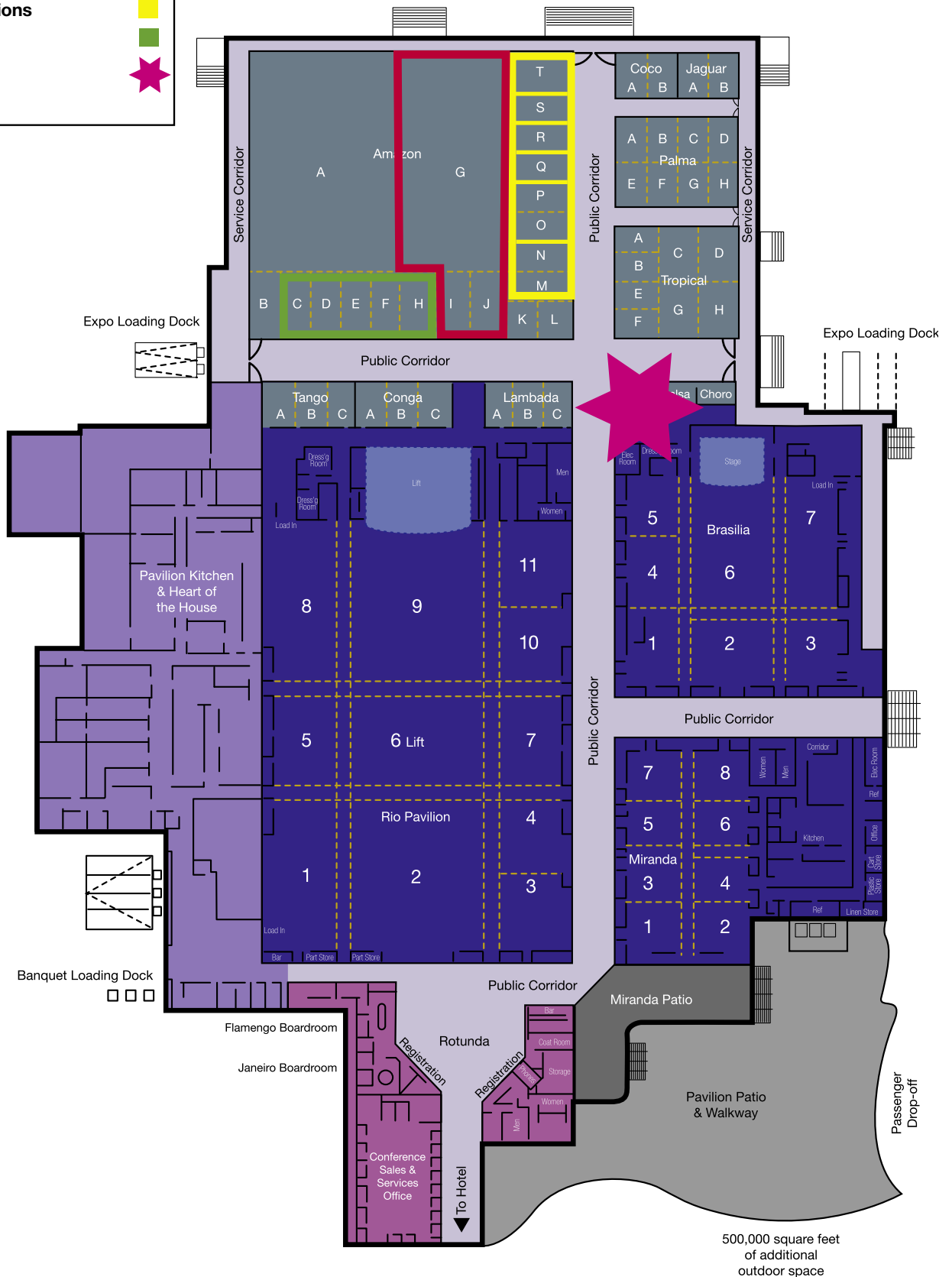*Note: No food or drink is allowed in the computer lab. Badges must be worn at all times during class and labs.
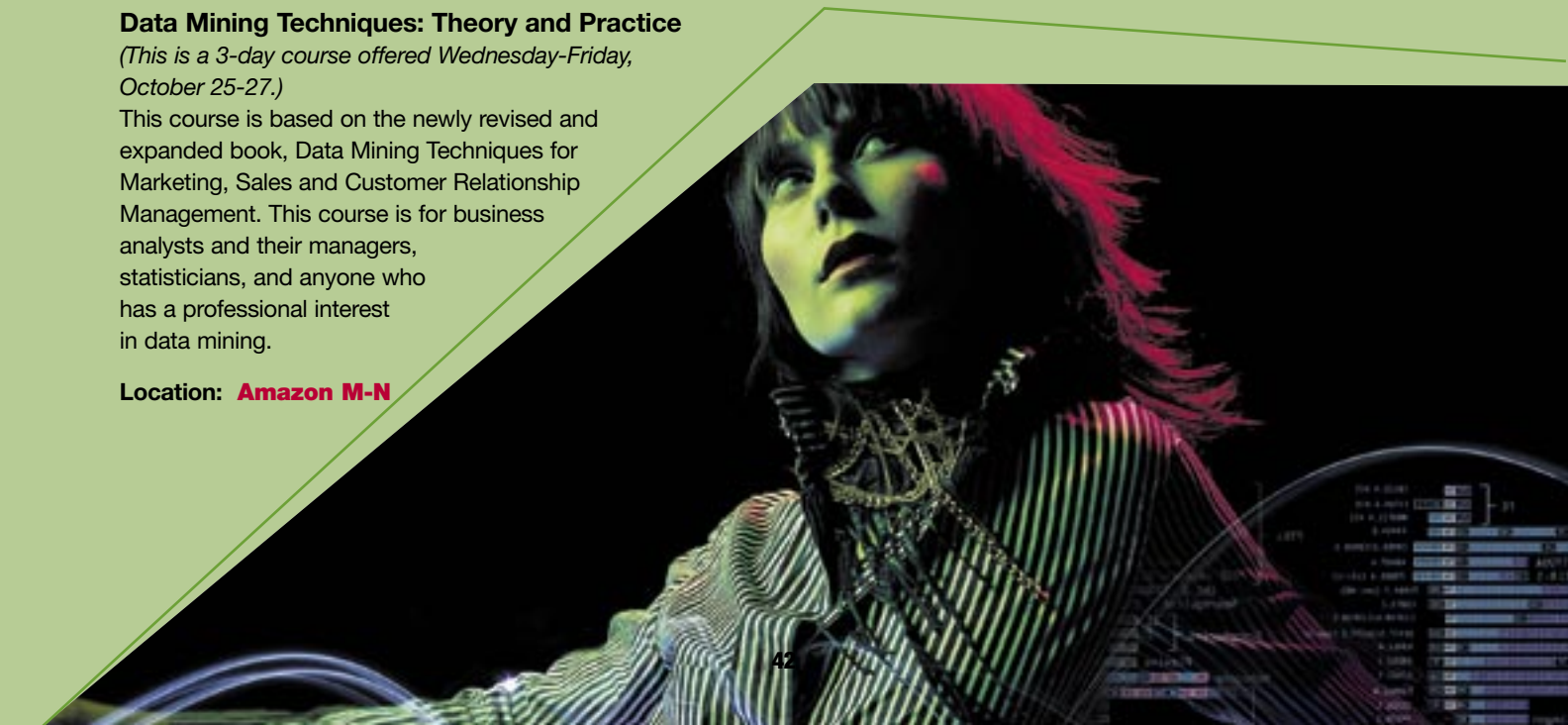
Map of M2006 Post-Conference Training at the Rio Hotel Conference Center

Lunch and Breaks

Training Sessions

PC Lab

Registration

Service Corridor

Public Corridor

Service Corridor

Amazon

A G

T
S
R
Q
P
O
N
M

Coco
A | B

Jaguar
A | B

Palma
A | B | C | D
E | F | G | H

Tropical
A
B | C | D
E
F | G | H

B C D E F H I J

K L

Expo Loading Dock

Expo Loading Dock

Public Corridor

Tango
A | B | C

Conga
A | B | C

Lambada
A | B | C

Salsa | Choro

Elec Room

Dress'g Room

Stage

Load In

Brasilia

5 7

4 6

1 2 3

Dress'g Room

Lift

Men

Women

Dress'g Room

Load In

11

8 9

10

Pavilion Kitchen
& Heart of
the House

5 6 Lift 7

Rio Pavilion

4

1 2 3

Load In

Bar | Part Store | Part Store

Public Corridor

Public Corridor

Corridor

Elec Room

7 8

5 6

Miranda

3 4

1 2

Women | Men

Kitchen

Office
Part Store
Plastic Store

Ref

Ref | Linen Store

Banquet Loading Dock

Public Corridor

Miranda Patio

Flamengo Boardroom

Janeiro Boardroom

Registration

Rotunda

Registration

Bar

Coat Room

Storage

Men | Women

**Pavilion Patio
& Walkway**

Passenger Drop-off

Conference
Sales &
Services
Office

To Hotel

500,000 square feet
of additional
outdoor space

Come see us at our new location next year!

Caesars Palace
October 1-2, 2007

M2007
10th Annual
Data Mining
Conference

Knowledge that makes a **difference**