

Success Stories in Data/Text Mining^[1]

Christophe Giraud-Carrier
Department of Computer Science
Brigham Young University

Abstract

This document presents a collection of successful implementations of Data/Text Mining. It consists mainly of excerpts from various sources, including vendors' literature and web sites, general data/text mining web sites, and proceedings of relevant conferences. It is organized by domains of activity.

Some statistics about this document:

Number of stories reported: 146

Insurances: 12, Retail and Direct Marketing: 22, Telecommunications: 10, Banks and Financial Institutions: 32, Health Care: 20, Public Transport (Air, Rail, etc.): 7, Pharmaceutical Industry and Bioinformatics: 7, Manufacturing and Production: 13, Miscellaneous: 23

Number of DM tools/vendors referenced: 30

Clementine (SPSS), Intelligent Miner (IBM), DD Series (DataDistilleries), Enterprise Miner (SAS), DataScope (Cygron), digiMine Services, WizRule/WizWhy (WizSoft), Data Miners, Quadstone System (Quadstone), Recon (Lockheed Martin), KnowledgeSEEKER/KnowledgeSTUDIO (Angoss), S-PLUS (Insightful), PolyAnalyst (Megaputer Intelligence), Cognos, Exclusive Ore, Business Objects, CART/MARS (Salford Systems), NeuroShell (Ward Systems Group), Z Solutions, CleverPath Predictive Analysis Server, Teradata Warehouse Miner, The Neural Network Toolbox (MATLAB), KXEN Analytic Framework, PolyVista Discovery Technology, E.piphany's product suite, Model Builder (Fair Isaac), ModelMAX, ScorXPRESS (ASA)

Insurances

Liverpool Victoria

- « As the UK's largest friendly society, Liverpool Victoria...now services over 2 million customers and provides a range of financial services including savings, investments and insurance. Today's financial services market is highly competitive and well saturated which...provide serious challenges for those companies committed to building their brands through individual service....Quadstone's technology was an ideal fit with the requirements of the database and market analysis team. Early results, such as increasing the response rate from a campaign by 35%, underlined the value of bringing in a tool that allowed marketers to take control of the customer profiling, segmentation and modeling process. One of the key commercial issues for any organization in the financial sector is customer retention, and recently Liverpool Victoria has been working on initiatives to improve their retention rates. Claire Davies, Senior Database Marketing Analyst, was given the task of helping to reduce the attrition rate of the motor insurance policy holders. A plan was developed to build a model that could be applied on a monthly basis to identify those who were most likely to switch to another provider with policies that were due to expire. A primary task Davies faced was to understand what data was available and how that could be used to help solve her business problem. Because Davies understood the business processes, she knew that some of the available data would need to be excluded since it didn't accurately reflect customer behavior. "Like every database marketing group we can't always have access to all the data we might like. However,

we are making use of the data we do have to produce something of real benefit," comments Davies....The work that Davies has done has helped confirm many of the instincts that the marketing department already had about customer behavior. But, more importantly Liverpool Victoria can now identify the combination of factors that most strongly affect behavior. »

Winterthur Insurance

- « As a combination of neural networks, the fully developed Clementine^[2] model correctly predicted who would cancel their policies from 60 to 70 percent of the time – proof of how predictive analysis can effect profits. First, Clementine produced a score for each customer, indicating the likelihood of cancellation. Then, Winterthur used each score to order the data. This resulted in 75 percent of these cancelling customers appearing within the first 15 percent of the dataset. Armed with the invaluable understanding of which customers were most likely to cancel, Winterthur was able to focus more readily on reducing customer churn and retaining profitable customers. »

ALKA^[1]

- « ALKA (www.alka.dk) is one of the five leading insurance companies in Denmark. To differentiate itself in a highly competitive market, the company aimed to optimize its expensive claim handling processes and offer better service to customers. ALKA sought to enable “fast-track claim handling” while still sifting out the fraudulent cases from its high claim volumes. Industry experts estimate that on average 10 percent of all claims are fraudulent, whereas usually only 0.5 percent to 1 percent of fraud is actually being detected. SPSS' software now enables ALKA to automatically determine during the claim notification process whether an incoming claim qualifies for immediate approval or needs to be further examined by experts. Hence, the company is now able to settle 8 out of 10 claims immediately within 24 hours after the notification, reducing claim cycle times and providing optimal customer service. »

Independence Blue Cross

- « During the last 18 months [from Sept. 2002], IBC's corporate and financial investigations department has referred 62 cases of suspected fraud or abusive practices to federal, state or local law enforcement or state regulatory authorities. During the same period, as a result of IBC referrals, 23 individuals or business entities were indicted on healthcare fraud charges and 14 people were convicted and sentenced. Edward J. Litchko, director of IBC's corporate and financial investigations department, said the company uses a sophisticated software data mining tool to analyze all claims submitted by medical providers and pharmacies and compare them against member enrolment data and overall provider information. Any trends, patterns or aberrant billing practices are targeted for an in-depth audit or investigation. IBC said it has recovered more than \$11.4 million from a variety of sources, including court-ordered restitution, returned medical and pharmacy claims payments, stopped claim checks, and providers who received payments to which they were not entitled. An additional \$6.8 million has been targeted for recovery and is being pursued. »

Spaarbeleg

- « To interact with their customers Spaarbeleg (a subsidiary of the AEGON Group, one of the world's five largest listed life insurance groups) uses a multi-channel approach, including mass media, direct mail, franchisees, independent brokers, Interactive Voice Response units and a sophisticated call center....The primary challenge was to get more sales out of the annually 2 million inbound customer calls....Until now inbound calls were limited to just handling complaints, transferring funds between the products or answering questions. Clearly, the inbound calls need to be transformed; from service calls into sales opportunities. The company

wanted to adapt its call center in such a way that the representatives are automatically notified of a customer's possible interest in a product, without being interrupted in their service activities or being forced into a sales conversation....To turn their service center into a profit center, Spaarbeleg selected DD Series, DataDistilleries' analytical Customer Relationship Management (aCRM)^[3] software. DD Series turns customer data into customer knowledge..DD/Dsire was implemented to provide customers with personalized product offers in real time.....Shortly after implementing DD Series, Spaarbeleg's CRM activities showed a dramatically improved performance: Up to ten million EUROS more sales and up to millions of EUROS of profits per year on 2 million calls...[and] An increase of the conversion rate by a factor of 100. »

Empire Blue Cross and Blue Shield of New York

- « In searching for fraudulent claims, analysts must look for unusual associations, anomalies, or outlying patterns in the data... By comparing the expected with the actual, large deviations can be found that can be more thoroughly investigated. For example, Empire Blue Cross and Blue Shield of New York used a database to compare the number of bronchoscopies reportedly performed by otolaryngologists. Newly reported numbers were compared to the existing numbers gathered from the insurer's 4.1 million members. The number of bronchoscopies from one otolaryngologist was well above average, and further research confirmed that the physician was submitting false bills...Empire estimates it saved \$4 million in 1997 alone by using data mining for fraud detection. »

FBTO

- « FBTO's [a subsidiary of the Achmea Group, one of Holland's largest insurance groups] primary challenge was to significantly increase the efficiency of marketing-initiated outbound customer contacts. FBTO set a very clear objective: increase the conversion rate of direct mail contacts with customers by 50%....FBTO selected DataDistilleries' open analytical CRM software for its ability to target customers highly effectively for outbound marketing activities....DD Series identifies those customers or prospects of FBTO who have a better-than-average chance of conversion for a financial product. Moreover, the software can discover customers' channel preferences, and the analyses include up-to-the minute customer data. This enables FBTO to offer the right product to the right customer through the right channel at the right time....[Benefits are] An increase of the conversion rate of direct mail contacts by 50%...Budget savings on channel costs per year...[and] A significant increase in customer satisfaction. »

Medical Benefits Fund of Australia

- « Chez MBF (Medical Benefits Fund of Australia), la première assurance de santé privée australienne, la fraude est estimée entre 1,3 et 2,3% des remboursements effectués. Soit 3,48 milliards de francs [français] en 1999....Celui-ci a recensé trois types de fraude émanant des patients ou des médecins : l'utilisation des droits d'autres allocataires alors qu'ils sont épuisés; la facturation de soins plus coûteux que ceux qui ont été effectivement administrés et celle de soins non exécutés....MBF travaille avec des spécialistes qui l'informent sur les genres de services ou de soins qui ne seraient généralement pas prescrits à tel type de malade en raison d'un autre traitement suivi, de son âge ou de sa situation. L'assureur cherche ainsi à détecter toutes les anomalies concernant les montants pour un patient, le nombre de traitements (radiographies, notamment) effectués par un praticien, la distance géographique entre lui et le patient, etc....En six mois, la solution de détection de fraude de SAS mise en place dans les services dentaire et optique a permis à l'assureur d'économiser la somme de 2,9 millions de francs [français], soit 200% de retour sur investissement en matériel et logiciels. La solution devrait ensuite être étendue à la branche hôpital. Cela devrait générer une économie de 16,3

millions de francs [français] par an. »

Amicon

- « Amicon's objective in using DataDistilleries is to increase the conversion on all of its health insurance products, such as national and private health insurance, travel, funeral, and study insurance, by providing better customized product offers based in a thorough understanding of individual customer behavior and needs. While Amicon already uses traditional statistical tools for complex modeling and reporting, DataDistilleries' advanced analytics software strengthens the marketing department by adding action-driven predictive analytics which result in better targeted and more frequent marketing campaigns. Using DataDistilleries, Amicon can build customer behavior models in just two hours, instead of two weeks, and conduct much more frequent and smaller campaigns to specific groups of customers. Amicon can also calculate the projected profitability of campaigns in advance, allowing them to plan marketing activity strategically, costing out expenditure against revenue...."The DataDistilleries software was put to immediate use, supporting our campaign calendar for Autumn this year [2002], and quickly demonstrated significant results: an increase of over 50% on the cross-sell conversion rate of our campaigns," says Henk Ramautar, Commercial Director of Amicon. »

Isapre Cruz Blanca

- « Isapre Cruz Blanca is the second-largest private healthcare insurance provider in Chile, serving over 500,000 citizens....Insurance is a highly regulated industry in Chile, and certain regulations create openings for fraudulent medical claims to be submitted and paid. Our goal was to dramatically reduce fraudulent disability claims to control costs and increase profitability....Using data from thousands of past claims, some of which were known to be fraudulent, Cruz Blanca used CA's CleverPath Predictive Analysis Server^[4] (which includes technologies formerly available as Neugents) to create a "Licenses Neugent" to detect new license claims that matched known fraudulent patterns...these suspicious claims were then presented to the audit group for closer examination. Also, the Licenses Neugent identified new types of claims that looked unusual....In many cases, these unusual claims revealed new types of fraud....We are accomplishing a much higher rate of disabilities fraud detection and have reduced costs by an average of four percent of the amount spent on each policyholder. We expect to save approximately \$2 to \$3 million dollars per year. »

The Hartford Steam Boiler Inspection & Insurance Co. (HSB)

- « HSB employs a range of risk and data analysis methods to manage underwriting and loss prevention strategies...We use a suite of statistical and analytical tools such as Statistica, Crystal Ball and @Risk, as well as internally developed risk simulation models. While all of these tools fill specific needs, we have searched for affordable, relatively easy-to-use data mining capability. DataScope^[5] fills this need in two important respects. It provides a robust set of tools that permit us to quickly explore data relationships, and the program's excellent graphics and straightforward interface permit us to perform real-time "what-if" analysis with clients. »

Major Insurance Company

- « A major insurance company is concerned with finding patterns in death claims, which occur within and shortly after the two years contestability period. During this time frame, the insurance company has the right to contest benefit payments. Therefore, it is imperative that fraudulent claims be detected in a timely fashion. Once the preliminary step was taken with a traditional auditing tool, WizRule^[6] was used to mine the data. Each cluster of interest was analyzed by WizRule for patterns in the data and for deviations. Since WizRule finds all the rules and relationships in the data without requiring pre-knowledge of the auditor, interesting

and unexpected results were obtained. Among the unusual patterns detected by WizRule were cases where the death occurred in a state other than either the state of insurance or the state the decedent lived in during the effect of the policy. In some cases, a pattern of foreign issued death certificates pointed to potential fraud. These cases are currently under investigation by the appropriate agencies. »

Retail and Direct Marketing

J. Sainsbury's

- « Quadstone began its relationship with Sainsbury's back in 1996, following the launch of their loyalty card program, Reward Card. Sainsbury's recognized that in order to deliver immediate value from the data produced by the scheme they needed to supplement their in-house analysis skills with specialist support. Quadstone consulting services provided the right fit between the business users and IT groups, and started working with Sainsbury's Relationship Marketing department to provide better customer-focused analysis of the Reward Card program. Following several successful early marketing wins that resulted from analyzing the Reward Card data, Sainsbury's purchased and installed Quadstone's technology in 1997. Quadstone was able to address Sainsbury's need for a very quick turnaround of various types of analytical query to answer specific questions concerning the behavior of the Reward Card customer base. It also provided the ability for more exploratory value-discovery through mining and visualizing the huge volumes of transaction data generated. »

Safeway UK

- « Intelligent Miner^[7] discovered that a particular cheese product, ranked 209th in sales, was frequently purchased by the top-spending 25 percent of customers – the last clientele Safeway UK would want to disappoint. Under conventional analytical principles, the product would have been “de-listed” ; in actual fact, the item was quite important. »

Marks & Spencer

- « Marks & Spencer is fondly known as Britain's “most trusted retailer.” Yet in recent years the company has discovered that you cannot simply live on your reputation. Marks & Spence has had to become more “customer facing” to create a platform for improved profitability...Data from the charge card system, combined with external sources such as census, demographic and national panel data, tells Marsk & Spencer a great deal about its customers. “We have at least 80 explanatory variables for every household in the UK, rising to more than 300 for any customer holding a charge card,” says Bond....By applying analytical techniques such as cluster and discriminate analysis and data mining, Marks & Spencer has identified 11 core customer segments (which are further subdivided). The analysis helps in a variety of ways, from corporate branding to operational decision making within the business units. It helps the company ensure that the products in a particular store are the ones customers want. Whereas in the past stores were stocked according to square footage, now they are increasingly supplied according to a detailed profile analysis of their customers....Careful analysis of customer behavior also facilitates more effective marketing activity. For example, the shopping activities of the 11 segments over the Christmas 2001 trading period were found to vary. By identifying who shops and when – for example, older customers tend to shop early to avoid crowds, while younger men leave things until the last minute – the company was able to align its marketing activity and product availability accordingly. The ability to do this was one of many actions the company took to help drive up the successful Christmas trading figures. »

Plow & Hearth

- « Plow & Hearth was already well established as a distributor of country style gifts when they became part of the 1-800-FLOWERS family in 1998. Founded in 1981 as small store in rural Virginia, Peter Rice had his mind set on a much bigger catalog business. Over the past 20 years, Plow & Hearth has grown and changed, turning a single retail outlet into a thriving multi-channel operation that includes a catalog, a call center and the internet. After reaching a circulation of 15 to 20 million in 1994, the next stage in Plow & Hearth's growth wasn't increasing the quantity of mailings but improving the efficiency and effectiveness of the books it mailed. Plow & Hearth used ModelMAX®^[8] to concentrate on the half million 12 month buyers in its 1.5 million name house file. They developed models to take better advantage of co-op databases and other prospecting efforts. By developing up-sell and cross-sell models using ModelMAX®, Plow & Hearth developed rudimentary "phone specials" for the call center agents. »

Williams-Sonoma

- « Using SAS's data mining software, Enterprise Miner, along with a suite of other SAS CRM applications, Williams-Sonoma has created a new marketing system that models and explores customer data from more than 30 million households, helping the retailer create a **personalized and cohesive shopping experience across multiple channels and multiple brands...** "With SAS, we know the most effective and efficient ways to contact our customers, as well as how often and when to market to them," says Dietrich [Manager of Statistical Development]. He uses Enterprise Miner to separate millions of customers into dozens of different segments and is able to accurately predict how each segment will behave. "We build hundreds of different customer models, and we're able to use that information to predict who is most likely to respond to a specific campaign or a specific mailing," says Dietrich..."With this information, we can specifically target the more responsive groups of customers..." After all, for a company that mails nearly 300 million catalogs a year, even small increases on return for each campaign can make a significant impact on bottom-line profits. "As a solution for predicting and targeting marketing returns, SAS is pretty easy to justify," says Holbrow-Martin [Director of Consumer Information Management]. **"Every tenth of a percent that we improve our targeted marketing efforts translates into millions of dollars in savings."**...In addition to its successful marketing programs, Williams-Sonoma also uses SAS in the process of determining strong regions for new retail stores. "We're leveraging customer information from SAS to understand which geographies have customers with high affinities to our products," says Holbrow-Martin. The company opened more than 35 Pottery Barn Kids stores in 2002. »

Thomas Cook

- « The Thomas Cook group has roots dating back to the mid 1800's, so you can bet it has amassed a huge amount of information about its customers. With a rich warehouse of data on more than 5 million customers, the UK leisure travel agency was missing one critical piece of the proverbial puzzle: the ability to fully utilize that data..."We had all this valuable data from different touchpoints and couldn't fully exploit it," says Paul de Laat, director of customer services and retention with thomascook.com. "...our method of analysing data was extremely inefficient. It could take us weeks to analyze customers according to several different criteria – such as age, past trip history, and touchpoint preference – at the same time..."Customer profile searches were conducted by an outside agency with turnaround time of about a week. A series of follow-up queries could stretch the process out for several weeks to even a month..."Bu we couldn't look at several different characteristics of individuals, look for people with similar characteristics, and then put them all together based on detailed customer profiles rather than a product or channel view..." The company chose E.piphany E.5 to kickstart its CRM effort....Using E.piphany Campaign Management, thomascook.com is developing highly targeted acquisition and retention campaigns based on this comprehensive view of its

customers. "For example, we can look at the profiles of people who've booked four-star hotels in Greece and compare their profiles with our database of 5 million retail customers," de Laat says. "We look at what other products are being bought by people who buy four-star hotel packages in Greece and then send, via e-mail, a special offer on those products. E.piphany helps us do this in-house in a couple of hours as opposed to taking three weeks by a service bureau before."...By merging customer data into a centralized system with analytic capabilities, thomascook.com is able to predict customer needs and bring customers back for repeat business. That ability to cross sell successfully, de Laat says, is going to generate the biggest return on investment. »

Argos, UK

- « Retailer Argos has installed data mining software to reduce losses from internal fraud and 'shrinkage'. The company is due to complete an initial rollout of the project at the end of March [2005].... Argos security solutions manager Adrian Sherry says the company needed more accurate measures for tracking cash and stock through its operations, to reduce losses. He says the new strategy will tackle such losses by analysing electronic point-of-sale data....Argos did not know how much it was losing each year from shrinkage - the unexplained disappearance of stock - so decided to invest in datamining software from vendor IntelliQ to provide its loss prevention department with more accurate data for targeting failing stores. 'We wanted to provide more value to the business,' he said. 'Before, more often than not, the high-risk stores we prioritised were dependent on stock and cash audit resolution from six-monthly figures.' The new software looks for abnormal behaviour at the till, such as fraudulent double-discounting, granting unnecessary refunds, or administrative errors such as not scanning the second product in a 'buy-one-get-one-free' offer. Daily reports are sent direct to the loss prevention officer via email, allowing better planning of which stores to target for investigation each day. 'The software can pay for itself with the losses we discover or prevent, while other departments like marketing and finance also gain benefits from its findings,' said Sherry. »

UNICEF Germany

- « We developed models that enabled UNICEF Germany to identify optimal target groups, thus increase its direct mail response by up to 80 percent...What's more, it showed a return on investment of more than 65 percent. »

CG2 Direct

- « CG2 Direct out of Wilmington, DE is a full service marketing services firm that has well known clients in banking, non-profit, and institutions of higher learning. CG2 is experienced at delivering turn-key solutions, which often include sophisticated analytics. In fact, they are so confident in their work that they frequently engage in 'performance based' projects. In other words, CG2 pays for up-front costs such as list rental, printing, and postage on a contingency basis. If the program makes money, CG2 makes money. One such project was a credit card acquisition program for Provident Bank. CG2 used ASA's predictive and descriptive modeling tool, ModelMAX® to manage a comprehensive program with Classic, Gold, and Platinum cards; pre-approved and by-invitation-only applications, and direct mail and telemarketing. ModelMAX® was used to perform custom demographic-based clustering and segmentation, and to create a triple filter that screened individuals for response, activation, and balance. As usual, CG2 was on target. An initial mailing of 1.5 million households led to additional phases and as many as 10-15 million pieces mailed. Provident's activation rates were 1.5 times the industry average and the average balance per active account was 40% higher than the industry average. »

Direct Wines

- « Direct Wines has grown significantly over the last few years, even faster than the market for wine, and now sends out 500,000 mailings a month to increase the size of their membership base...As the company grew, it started to segment its customers rather than blanket mail the whole database....Direct Wines worked with SPSS in order to reach customers with more targeted specialised offers. The use of sophisticated data mining software helped them to build on their existing database and personalise customer offers more efficiently....With SPSS, we can use our own data as well as external research to determine sales pattern so that each mailing we do is backed by evidence of previous success. »

Allrecipes.com

- « As a high-traffic Web site, Allrecipes.com must remain in tune with its visitor behavior in order to keep the content compelling, easy to navigate and reliable....Accurate and timely usage of information is crucial...to evaluate the effectiveness of new services and to quickly identify and resolve any problems or bottlenecks on the site....Allrecipes.com initially attempted to tackle the challenge with standard Web log analysis software packages. We quickly found that these software packages were processor-intensive and produced unsatisfactory results....Allrecipes.com identified [digiMine^{\[9\]}](#) as the best solution for us because it provides powerful analytics through a cost-effective service. DigiMine builds a secure data warehouse and analyzes user behavior, providing us with a complete data asset and decision-support reports....we are now able to be proactive with our content management....we are able to be more responsive to our visitors' needs, which is most important to us. »

Woodcraft Supply Corp

- « Woodcraft Supply Corp. is one of the nation's oldest and largest suppliers of quality woodworking tools and accessories. Located in Parkersburg, West Virginia, they market through catalogs, internet sites, retail outlets, franchise stores, and Woodworkers Clubs®. Woodcraft annually distributes more than three million catalogs featuring nearly 8,000 items to all 50 states and 117 foreign countries. Carefully managing costs and improving the ROI of each promotion was important to Woodcraft, as is the case for most catalogers. Woodcraft has used ModelMAX® since the mid-90s, initially focusing on improving the cost efficiency of catalog mailings. ModelMAX® enabled them to reduce the number of pieces mailed by 13% and increase net profits by over 45%. As great as these results were, Woodcraft continued to look for additional ways to improve their targeting efforts and ultimately, their bottom line. Rather than use a single previous mailing to predict a single future one, Woodcraft decided to try an innovative approach. By combining the results of three previous mailings to predict a single future mailing, Woodcraft uncovered even higher response rates, and of course, profits. ModelMAX® accurately predicted a 47% response rate from the top segment of the combination model. »

Envision EMI, Inc.

- « Envision EMI, Inc. manages and markets programs designed to further the goals of high-achieving students....Envision tries to establish personal relationships with their contacts in each high school that they target – all 33,000 of them....To intelligently extract data from their legacy systems, Envision's marketers had to program their queries and send them to a vendor for hard coding in SQL. That process took up to six weeks to complete....Envision decided upon E.piphany Insight to get a total view of its client base to develop intelligent marketing strategies. With the E.piphany system's sophisticated data mining, Envision quickly identifies its best and worst customers. With one click, Envision's marketers can move into campaign management....As a result of deploying the E.piphany system, Envision has begun to change the way it plans and executes marketing campaigns. With the ability to easily model and

explore various what-ifs, the company has become more strategic and devotes more time to campaign design. The 20 people who handle marketing campaigns no longer have to function as statisticians, and data mining discoveries quickly spread throughout the organization, stimulating collaboration. DeCagna [Envision's chief officer of the global marketplace] says that Envision can now develop sophisticated campaigns with many segments and treatments in as little as one day. »

CustomerLinx

- « Traditionally, telemarketing businesses have billed their clients based on the number of hours they spend on a campaign. Given that traditional campaigns could be considered successful with a 1-2% response rate, just a small percentage increase in effectiveness can improve profitability dramatically. CustomerLinx's bold business program uses "in-campaign modeling" which gives the company the confidence to commit to a new pricing structure. Clients enrolled in this program are now only charged when an actual sale is made. The CustomerLinx approach is based on analyzing customer behavior patterns quickly so campaign managers can effectively target those most likely to buy while the campaign is still running. This "pay-per-success" program is a radical new approach that brings unprecedented levels of efficiency and effectiveness to telemarketing campaigns. Quadstone's technology enables CustomerLinx business professionals to understand customer behavior enough to refine and redirect resources on the fly....By using in-campaign modeling, CustomerLinx has achieved excellent results. For example, in one campaign CustomerLinx increased response rates from 0.8% to almost 4.0%. As Wyatt explains "Quadstone's, Decisionhouse, not only saves money, but makes money." »

Cécile Co., Japan

- « "Swimsuits sell well by mail order in the cold region, but jackets do not." This seems to defy common sense, but it is a fact that Cécile Co. executives discovered while using SAS software. Detecting sales trends can be difficult. But when analysts let data speak for itself, it can yield surprising and lucrative discoveries. Cécile, one Japan's largest mail-order companies, is using Enterprise Miner^[10] for a new generation of business data analysis and data mining to become a leader in one-to-one marketing....In the past, Cécile executives analyzed customer trends by repeatedly refining statistical models. They used their database to periodically test new models and new hypotheses. It often took managers weeks or months to make decisions about catalog content. But with Enterprise Miner, Cécile executives can discern patterns in customer-spending data in one-third the time. The old, time-consuming process of forming and testing analytical models has given way to discovering trends in customer spending that were not even imagined before. And this new information allows for much more sophisticated marketing....Cécile's customer database is the largest in the mail-order business. Its 19 million customers can be segmented into more than 100 groups based on their purchasing habits, trends and tastes. Each group can now receive targeted catalogs that contain the latest seasonable fashions. Cécile also uses the data to conduct direct-mail campaigns and other marketing activities that offer better cost performance. »

Miami Herald Publishing Co.

- « The database marketing group contacted all of the households in its market area several times per year offering incentives for short-term subscriptions, resulting in a high cost per new subscriber. To increase profitability, the group used techniques including segmentation and customer profiling to find households more likely to be stable, profitable customers. In turn, the group focused marketing efforts in that direction...By targeting individual households, the Herald offers advertisers both traditional and more creative marketing avenues, from targeted inserts to joint direct mail campaigns with advertisers. »

Jubii, Denmark

- « Jubii is achieving a high level of personalization with the help of Clementine, SPSS' data mining workbench...Beginning with a relatively small number of loyal users, the project assigned profiles to each of them based on the pages they viewed and the time of day they visited. These profiles, which are updated upon each new visit, are then fed into data models that determine which banner ad is most appropriate for the visitor. The initial project was so successful – ad click-through rates increased by 30 to 50 percent – that it has recently been expanded to cover the entire portal. »

Eddie Bauer

- « Eddie Bauer uses predictive modeling to decide who receives specialized mailings and catalogs. For example, each year Eddie Bauer features an outerwear special, and – thanks to data mining – it can determine which customers are most likely to buy. Data mining allows Eddie Bauer to determine seasonal buying habits as well. Then the company can identify people with similar characteristics who don't normally buy outerwear and target them with mailings to bring them into the store or encourage them to buy from the catalog....customer spending has increased by 30 to 50 percent annually. »

Fingerhut

- « Fingerhut is the second-largest cataloger in the world selling a road range of products and services directly to the customer through catalogs, direct marketing and the Internet....With sales of nearly \$1.8 billion and data on 30 million households, we needed to find a method that would help us deliver the right catalog with the right products to the right customers....The Quadstone System^[11] is primarily being used by Fingerhut to analyze online and direct marketing campaigns. It is also being employed to make predictions of customer behavior in areas such as response to promotions, order size, return of orders, pre-shipment cancellation of orders, lifetime value of customers and saturation promotions....Early tests with the Quadstone System returned results showing potential improvements of a few percentage points on campaign effectiveness improvements that could amount to hundreds of thousands of dollars in savings. »
- « Fingerhut Companies, a Minneapolis-based direct-mail catalog company using SAS' server-based Enterprise Mining...is sifting through a database of 10 million to 12 million current customers to find which are most likely to buy products from one of the company's many catalogs. Fingerhut, which has 9,500 employees and mails 130 different catalogs each year, is among the true believers in data mining: All catalog mailings, credit-granting decisions, and inventory-stocking decisions are based on it, says Andy Johnson, Fingerhut's senior vice president of marketing. Fingerhut wants to find out which customers it could profitably mail catalogs. It recently used data mining to study past purchases of customers who had changed residences to see if they had preferences. Data mining showed those customers were three times more likely to buy items such as tables, fax machines, phones and decorative products, but that they were not more likely to purchase high-end consumer electronics, Johnson says. Fingerhut used that information to create a special catalog it mailed only to those customers who had recently moved. »

The Vermont Country Store.

- « For years, the store's mailings had been based on customers' buying histories: how recently and how frequently they purchased and how much money they spent. The "recency-frequency-monetary targeting" (RFM), is quite common in the catalog industry...But while RFM is a fairly reliable modeling method, models are limited only to those three variables....During the evaluation process, a database analyst at The Vermont Country Store

with no statistical training used Enterprise Miner to create a model that predicted a 2.83 percent increase in results. “We had better predictions available at the time, but we are traditionally a very conservative company, so we used the lowest one we had,” says McCarthy [Manager of Statistical Services and Market Research]. “But when we applied that model and projected it out through the year, with our 20 mailings totaling 40 million catalogs, we came up with more than 1,000 percent return on investment. It was a real slam dunk.” With SAS firmly in place, McCarthy says that The Vermont Country Store’s predictive capabilities have increased exponentially. “While RFM targeting limited us to just three variables, Enterprise Miner allows us to evaluate up to two or three hundred variables, and we’re no longer limited to recency, frequency and monetary data,” she says....By using Enterprise Miner to group and analyze customers, the company can profile customers and determine their relative importance to the business as well as their likely shopping behaviors – and decide which customers should receive which catalogs. »

Experian

- « ...Experian, the world’s largest provider of consumer and business information, as implemented the KXEN Analytic Framework^[12] within its Decision Support & Analysis Group....”KXEN systems have greatly expanded our already powerful arsenal of analytical capabilities,” said Marc Fanelli, vice president of Experian’s Decision Support and Analysis Division. “KXEN’s ability to synthesize massive amounts of dynamic data quickly and efficiently allows us to spend more time providing consultation and recommendations, regarding contact strategy development, to our customers.”....Experian clients purchasing prospects information identified with the aid of KXEN will enjoy reduced modelling turn around-times, state of the art targeting technology, and web-enabled reporting of analysis results. »

Medium-sized software company

- « The marketing department of a medium sized software company was charged with developing a strategy for the introduction of a new product. A strategic decision was made to design and implement a telemarketing campaign to sell the new item to previous customers. The customer database consisted of 30,000 names, as well as several fields for each customer; such as, address, date of previous purchases, modules purchased, etc. First the marketing group randomly selected 400 customers. Of that sample, 50 cases had incorrect phone numbers, leaving a viable sample of 350. All were called and pitched the new product. Sales were recorded in 9 calls. The success rate with this method was 2.5%. WizWhy^[13] was then run on the same sample. The key prediction parameter (dependent variable) in the WizWhy analysis was whether or not the customer had purchased the new product. WizWhy automatically discovered all patterns and characteristics of the customers who were purchasers, as well as those who did not purchase. The full customer database (less the sample) was then subjected to the WizWhy prediction mode. WizWhy made a prediction of the conclusive probability of purchase for each customer in the population, based upon the profiles it had revealed in the sample. The database was sorted according to this conclusive probability. A telemarketing campaign was then targeted at the 1000 customers with the highest conclusive probability, i.e., those most likely to purchase. Of the 1000 customers called, 60 in fact purchased the new product. Using WizWhy to determine the profile of likely buyers increased the success rate from 2.5% to 6%! »

(Tele)Communications

Globo.com

- « Globo.com, one of Brazil's internet service providers, is using SPSS predictive analytics to improve customer service, acquire new customers and retain existing customers. Harnessing SPSS predictive analytics, Globo.com has acquired 200,000 new customers and reduced customer churn by 30 percent....the Internet service provider is now able to segment customers based on usage level and churn motives, enabling it to create loyalty strategies based on discounts, packaged plans and specified content for each user. SPSS for Windows also enables Globo.com to segment its customers by topic, allowing for up-sell opportunities and increased revenue. For instance, Globo.com can e-mail premium sports videos to customers to help them follow their favorite teams. "With SPSS predictive analytics, we have been able to better understand the motives of churn, allowing us to implement an effective retention strategy," said Jeremias Coral Klausner, Globo.com's Data Mining Coordinator. "We're now able to better focus on our business needs, and more importantly, our customer's Internet needs." »

One 2 One

- « One 2 One's marketing department used Business Objects^[14] to determine prioritisation of network roll-out during the company's development of UK wide coverage. Having identified a typical One 2 One customer type, the company used Business Objects to analyse the geo-demographic data and determine the most profitable potential areas of the country for One 2 One. The network was rolled out on that basis, with new cells being opened first in the most profitable potential areas. »
- « One of One 2 One's most successful promotions was targeted at customers who had set up their system to support roaming while abroad and had not yet made use of it. Caroline Moyle, Analyst Programmer at One 2 One, explains, "We used Set Analyzer to create a customer set based on customers who had made a call that could have been at an airport, then made no calls for a day or two, and then made another call at an airport. We made the assumption that whilst they were not making calls, they were abroad. That level of segmentation achieved by Set Analyzer was critical, we could not have made that mailing without it and it was extremely successful, generating the highest response we've ever had to a mailer." »

US West

- « After a major promotion for a new Business Voice Messaging system, the company experienced a rate of churn, in excess of 40%, a huge revenue drain on a project that was still doubling in sales on an annual basis....We looked for a solution with KnowledgeSEEKER^[15] by concentrating on the discovery of actionable variables....If you are looking to discover interactions between a lot of factors, KnowledgeSEEKER has the ability to identify interesting relationships you would never have seen using the logistic regression techniques....We got through the data quickly, and in a short period of time were able to narrow the list of potential causes of churn from 50 down to 3 actionable variables....The three causal factors identified by KnowledgeSEEKER if implemented, could have reduced the churn rate by 50%. »

British Telecommunications

- « The once peaceful telecommunications industry has turned cutthroat....To retain its customers, gain new customers and maximize sales, the company needed facts about exactly who was buying its products and services....During data analysis, the team employed Clementine to identify data quality issues, become familiar with the data and data distribution and eliminate data attributes not strongly associated with the purchase of Business Highway. Then, it measured the predictive strength of individual data attributes in relation to the customer's propensity to buy the product. For example, two-digit district codes, a geographic indicator, were clearly linked to response and purchase data. After the analyses, the team quickly built and tested a series of experimental models using Clementine's decision

trees....”Our deliverables to sales and marketing were lists of customers and charts showing why these were the customers they should speak to about the Business Highway product,... Even before completing the final models, we were able to surpass our original target – and increase the campaign response rate by 100 percent,” O’Brien said. »

ECtel

- « Ectel, a global provider of monitoring solutions and revenue assurance applications for wireless, wireline and next generation networks, has incorporated Clementine to help deliver real-time fraud detection solutions to more than 150 telecommunications companies around the world....The worldwide telecommunications industry loses more than \$12 billion dollars in revenue each year due to fraud...With the powerful combination of Clementine and FraudView, Ectel’s real-time fraud management system, Ectel customers can instantaneously detect and stop telecommunications fraud – resulting in significant money savings. To date, more than 150 blue-chip telecom service providers worldwide, including national carriers in Germany, China and the United States, have turned to Ectel’s FraudView to successfully inhibit fraud.” »

Verizon Wireless

- « To bring churn into control, Verizon decided to develop churn predictive models using data mining techniques. The models were used to analyse outputs to identify those sets of customers who could be looking at exiting from the company. It also used that predictive modelling to target customers with specific, relevant and timely offers. Simultaneously, the company decided to use data mining techniques to predict the right kind of plan for any subscriber based on his/her call usage pattern. Up until this point, data mining was basically an IT initiative in the company. In order to derive the best results, IT brought the idea to marketing team and presented it as a partnership project within the organisation....Verizon used data mining techniques to do predictive churn modeling (PCM) and predictive takers modeling (PTM). Simply put, out of the total targeted customers who would have high propensity to leave, Verizon had to use various techniques to arrive at a very specific small group, who would end up as actual takers of Verizon's offers. While PCM was used to find a smaller group than the overall target group, PTM was used by the company to target only those customers who were likely to stay. In the process, the company sent mailers to a relatively smaller number of people and saved on costs incurred over those customers who are less likely to stay. Verizon was able to reduce its churn from over 2 per cent to below 1.5 per cent using the data mining tools. Ms Kronic [Principal Data Mining Engineer] said Verizon was able to achieve this because the company had a critical, highly visible business problem on which it focused. The fact that a multi-disciplinary team with specific area experts was used was another important reason among others for the success of the project....Various business benefits that accrued to Verizon Wireless due to adoption of data mining techniques involved cost reduction, revenue increase and learning processes among others. The company's direct mail budget for "churner mailing" reduced by 60 per cent for highly targeted marketing campaigns. The revenue increase could be seen from thousands of contract renewals per month and increased usage by subscribers. »

Hutchison Telecommunications

- « Hutchison Telecommunications (Hong Kong) Ltd. has chosen Enterprise Miner from SAS to gain and sustain lasting competitive advantage through better business insights....”SAS Enterprise Miner helps us profile customer activity through the data mining process,” says Ronald To, system architect at Hutchison Telecom. “By analysing and reporting on our ‘customer view,’ every communications and marketing channel can be used more effectively through a detailed analysis of customers’ behavior and preferences, giving us the opportunity to exploit this knowledge and accurately drive our business forward.” »

MTN

- « MTN's analysis of its prepaid mobile phone customers seemed like a shot in the dark in the beginning. "When we began our analytical study of prepaid customers, there was no guarantee that we would achieve any real predictors at all, much less churn management. This was because of the lack of demographics and the size and scope of the data we we're using," says Marcelle Georgiev, customer knowledge analyst at MTN. But in the end, Georgiev's group achieved an impressive result that far surpassed expectations: "We achieved an 82 to 88 percent accuracy rate, depending on which cut of the data we used," she says....Mohan Namboodiri, data mining program manager at SAS South Africa, explains the situation: "For customers that are on contracts, churn is fairly straightforward – the contract ends and the customer goes away, for a number of reasons. Things move much faster in the prepaid market, but we studied churn in the contract market to give some clues. We wanted to see if the data could speak to us about churn....The study yielded more than a clue, it predicted some significant trends...."The whole issue of the peak week, a drop in peak week outgoing calls, and an increase in weekend calls, that was something that we discovered. The thing that makes it interesting is, MTN didn't have peak or off-peak rates then. It was just a lifestyle pattern we uncovered, as opposed to cost. At the time, MTN's customers had all prepaid a flat rate. Since then, MTN has revised its rates and now has off-peak rates and international rates. It will be very interesting to see what comes up now."...The way MTN views prepaid customers and the future of customer relations has been positively influenced by the data mining study done with SAS software...."Our work is changing perceptions in the company. People assume that prepaid customers want anonymity, will not be loyal and do not want any contact with the company, but that is not necessarily true. Once we realized how stable the higher-value groups are, we made a big effort to get them onto the loyalty scheme, which could even increase their revenue," says Georgiev. »

Large Wireless Provider

- « ...the largest wireless provider in its home country, was faced with competition from several other providers. One of the biggest challenges facing them was customer churn...the business users told us [Data Miners][\[16\]](#) to build separate models for high value customers and everyone else. The data told us to build separate models for recent customers and older customers... Listening gave us the insight to model four customer segments...These resulted not in a one-time improvement to lift, but in a sustained process improvement for their database marketing team. »

British Telecom

- « In the area of marketing, BT used KnowledgeSEEKER to help market a previously marketed product. They wanted to understand which customers should be approached. KnowledgeSEEKER was used to create a customer profile for a product similar to 'Friends & Family'. KnoweldgeSEEKER also found new geographical distribution information and other attributes which were very valuable. It was found that KnowledgeSEEKER dramatically reduced manual tasks... »

Bell Canada

- « We produce over 1,500 specific direct marketing lists annually in support of a wide variety of wireline, wireless, Internet and bundled telecommunications offers and channels....We were looking for ways to improve our target modeling capabilities in a competitive and results-driven environment....We can visually examine the strongest factors associated with a particular offer...quickly discover what factors help to "split" customers...discovering how different factors interact....The applications of KnowledgeSTUDIO[\[17\]](#) have been quite successful in

meeting our original business needs: we have reduced our modeling development time by 40 to 50 percent and have also seen an incremental improvement in our model targeting power of 5 to 10 percent based on this software. »

Banks and Financial Institutions

Banco Espírito Santo

- « The bank's commitment to customer service led its strategic marketing group to articulate a clear mission: implement analytical tools and adopt relationship-building techniques to precisely predict customer behavior and increase customer loyalty....Jorge Portugal and his team at Banco Espírito Santo (BES) devised a three-step, data mining process to maintain long-term relationships with their customers. First, they set out to identify the different categories of customers by analysing their transaction behaviors. Building highly detailed and dynamic models, or mathematical equations representing these behaviors, was the next step in their analytic process. And, after many iterations, Jorge and BES plotted an intervention strategy to satisfy their customers' needs-and keep them from going to the competition....The marketing gathered statistics on customer behavior and customer satisfaction from every possible source, including Web logs, transactions, demographic and third-party data...."SPSS empowered us to know ahead of time when a customer is at risk of leaving," he said. "Now we can take the appropriate action to keep them." With this system, the team was able to reduce customer asset erosion 15 to 20 percent by refining the BES product line and service offerings...after identifying and modeling the behavior of potentially migrating customers, they can take proper action to retain those customers....By spending time with important customers, BES is increasing bottom line profits by 10 to 20 percent this year [2001] alone. »

Lloyds TSB

- « Dan Golding, fraud prevention manager of Lloyds TSB Cards in the UK, says his company has identified and potentially prevented more than £2 million (US\$ 2.9 million) worth of fraudulent activity in a year's time using SAS' data mining software, Enterprise Miner. »

Associates Finance

- « DataLab USA has won the 2004 Bronze Award for Excellence from the National Center for Database Marketing (NCDM) on behalf of its work for Associates Finance Puerto Rico.... The direct mail program, a highly targeted database marketing customer acquisition campaign, reversed a trend of declining response rates and decreasing margins for Associates Finance in a small, fiercely competitive market of loan prospects in Puerto Rico, which has an overall population of 3.5 million. "Because Puerto Rico is an island nation, the market for loan prospects is extremely small, at about 600,000, so this was an enormous challenge for any direct marketing vendor," said Charmaine Llauger, Director of Marketing, at Associates Finance in San Juan.... DataLab USA implemented a novel two-tiered database marketing approach utilizing a newly developed predictive model and "re-mail penalty" segmentation technique to boost the declining results of the personal loan direct mail program. The innovative strategy, designed to reduced overall mail volumes while increasing mail frequency, increased program response rates by more than 10%. »

Barclaycard

- « The company had seen consistently good results from its adaptive control approach to account management strategies...But portfolio managers wanted to grow further – and faster. They wanted better returns on credit limit strategies and, in particular, they wanted to increase interest-earning balances....Barclaycard representatives had seen a demo of Fair, Isaac Model

Builder for Decision Trees, [\[18\]](#) a tool that uses historical data to assist in identifying optimal account management strategies. Barclaycard thought this just might provide the “kick start” they were looking for.....Fair, Isaac worked closely with Barclaycard at every step of the way....The total implementation time, from data extract to final strategy implementation, took only three months. After six months, evaluations of new strategies showed a significant impact on profitability, as can be seen in one of the strategies developed using Fair, Isaac Model Builder for Decision Trees. This strategy saw an 8% rise in net interest income after bad and doubtful debt, compared to Barclaycard’s existing champion strategy. Other strategies showed similar improvements. “While the software plays a vital role in the project, it is also important that as much human expertise as possible is used to guide and interpret the software.” Wiltshire [Customer value Management at Barclaycard] notes. “In addition to the bottom line benefits we are realizing, the process also helps us in understanding of the drivers of the strategies, and ultimately, our customers.” »

Keystone Financial

- « Keystone Financial (now part of M&T Bank) used ASA's predictive modeling and profiling tool, ModelMAX®, to grow and expand their home equity loan portfolio. ModelMAX® helped Keystone squeeze an additional \$1.6 million of net revenue from just three promotions. Because the results were so good and the program was so easy to maintain, Keystone continued the program for quite some time. But eventually results started to slack a bit simply because Keystone was continually 'fishing from the same pond'. Keystone's prospect file contained fewer and fewer high scoring candidates, credit bureau costs continued to rise, and delinquency and charge-offs were becoming more of a concern. Keystone wanted to keep offering home equity loans, but was not sure how to regain the previous levels of success using ModelMAX® analytics. ASA helped Keystone use the descriptive power of ModelMAX®, to perform customer profiling and modify their customer selection process....Results using ModelMAX® showed that six characteristics were key to properly selecting prospects. Applying this new filter allowed Keystone to eliminate 100,000 poor prospects, and second model eliminated another 10,000. The final tally? The new steps contributed an additional \$1.12 million toward the overall profitability of the project. Perhaps best of all was the cost to rejuvenate a program that at first glance looked as if it was past its prime--Keystone simply invested a little time. »

Morgan Stanley

- « Locked away in company databases and legacy systems were volumes of information about the 5 million accounts managed by the firm, but there was no clear way to gather and make sense of all that raw data and create a holistic view of clients....LoFrumento’s [executive director of CRM] team uses SAS for data warehousing and data mining in conjunction with SAS Marketing Automation and SAS Strategic Performance Management solutions....The result? Morgan Stanley now has a CRM data mart that captures all client information in one place, providing an image of each client that offers unparalleled precision and detail. With the information infrastructure in place, LoFrumento’s team can perform the analyses needed to bring a higher level of accuracy to every facet of Morgan Stanley’s business....A recent IRA rollover campaign is one example. Rather than mail to its entire customer base as it had done in the past, Morgan Stanley used a predictive model to select the appropriate audience of investors from the CRM data mart. The result was a 40 percent increase in the number of accounts opened during the same period the previous year. »

Provident Financial

- « Using Clementine, we can change our models as often as we need to. We can retrain the Clementine neural network and rule-induction engine whenever we detect a shift in fraud

patterns and profiles. Nearly 80 percent of the fraud discovered is identified using Clementine modeling... »

HSBC Bank plc

- « About 125 million credit card transactions are made each year by HSBC bank cardholders. Each transaction carries an income fee payable by the receiving bank to the cardholder's bank....We believed that incorrect fees were being received for some transactions....A daily transaction feed was sorted using SPSS software and within two weeks income shortfalls had been identified. Further manipulation of the data indicated substantial variances between the income received and the income that should have been received. We anticipate that recovery of this underpayment will generate in excess of one million pounds per annum. »

Fireman's Fund

- « Fireman's Fund competes for a share of the U.S. property and casualty market in a dozen lines of business, many heavily regulated at state and federal levels. To succeed, we need a clear understanding of what makes us profitable in order to make intelligent and timely decisions that are based on good information, allowing us to consistently meet customer expectations....SAS enables us to aggregate data from source applications and various data warehouses on a "sandbox" server for data fusion and analysis....With SAS, we can build models 20 to 30 percent faster than was possible before implementing this solution....The result is better use of staff resources, less duplication of efforts and significant time savings....and adjusters echo these sentiments. They are able to focus their activities with greater accuracy because we can quickly build statistical models that have highly relevant and useful predictions....With the models created by SAS Enterprise Miner, we can fully leverage the information stored in our data. We've been able to note suspicious correlations and flag anomalies for follow-up by specialized investigators, helping to root-out fraud and increase employee productivity. We've also used SAS to uncover premium fraud and identify cases that are candidates for subrogation (claims we service and pay, but for which others are ultimately responsible), reducing our claims liability. SAS models help us rank our customers so we can better meet their unique needs. For instance, we can move higher-claims customers into different programs while simultaneously focusing on retaining lower-claims clients. »

Beneficial National Bank

- « Beneficial National Bank (now part of Household Finance) is one of the largest private label credit card providers in the industry... a new merchant needed high volumes of mail-in applications processed. As many as 5,000 to 7,000 new applications were being received on a daily basis. Mail-ins had historically represented the highest risk for fraud....Fraud analysts identified 29 fraudulent applications in one day's worth of mail. Naturally, it took many hours of manual labor (not to mention years of experience) to accurately spot the fraudulent applications. ScorXPRESS®^[19] identified 28 of the 29 fraudulent applications, and recommended a severely restricted limit on the 29th. Additionally, ScorXPRESS® identified 18 more applications that were suspect but went undetected by the fraud analysts using traditional methods... the savings for this one client alone in reduced fraud and staffing costs came to \$1.3 to 1.5 million! »

Postbank

- « Postbank has more than 7 million account holders and conducts 5 million financial transactions a day, making it a leading financial services organization in the Dutch market....Postbank recently made the transition from being a product-focused organization to a customer-focused company. An important challenge arose for Postbank during this transition, which was to improve analysis procedures: the marketers wanted more frequent use of models

for campaigns, and to be able to modify them more quickly for immediate usage....Postbank selected DD Series, DataDistilleries' solution for analytical Customer Relationship Management (aCRM), to optimize its analysis procedures. The advanced Customer Behavior Modeling function in the software, analyzes the customer database of Postbank, and defines the optimal target group for direct mail campaigns....The analysis function lies hidden behind a user-friendly interface, enabling marketers to independently produce a model with the touch of a button....By implementing DD Series, Postbank was soon able to: Improve the conversion rate by 25%...Shorten the time-to-market from weeks to days...[and] Directly save a few million Euros in marketing costs. »

Credit Suisse

- « Competition in the financial industry is intense, and obtaining new customers is an expensive proposition....In 1997, Credit Suisse started the “Loyalty Based Management” program, with the primary goal of retaining profitable customers. They invested in a six-member “data mining team” that used the Clementine tool from SPSS Inc. to analyze a robust datawarehouse of its 2.5 million customers, each with more than 400 attributes. The analysis was used to identify potential leads among Credit Suisse's customers and intelligently market to them based on their individual preferences and histories....With Clementine, Credit Suisse can now identify customers, typically the top one percent, who are extremely likely to buy a service, thus increasing the opportunities for cross-selling and retaining customers. Detailed segmentation of its vast customer base allows Credit Suisse to develop targeted solutions for its customers....Credit Suisse used Clementine to analyze situations where customer interest in a service did not correlate with a purchase. Many times, customers did not have good enough credit and were subsequently refused the service. Improved models factored in credit as a criterion. As a result, the percentage of customers interested in purchasing a service but who were refused due to bad credit was reduced by almost half in subsequent campaigns....Dr. Nippe [head of data analysis/data mining] affirmed, “we recouped the total costs of the project within two years.”. »

Teikoku Databank

- « ...Teikoku Databank (TDB), Japan's largest corporate credit research company, has developed a statistical model to compute the likelihood that a prospective corporate borrower will go bankrupt within a year. Bu the SAS application does more than help shield lenders from high-risk credit exposure. It also automates and speeds up the credit approval process, alleviating a major challenge for financial institutions and other companies that deal with a large number of small accounts. It is powered by Enterprise Miner software.... »

Standard Life Bank

- « With SPSS data mining capabilities, Standard Life can address a growing number of projects enabling them to deliver better value to their customers. With the aid of SPSS, they were able to build a propensity model for the remortgage offer showing the types of clients attracted to this product....Using the propensity models, a remortgage mailing campaign was planned and executed by the Customer Data Analysis team in conjunction with Standard Life Bank. The mailing included a randomly selected control cell. The model allowed the bank to focus their efforts on the best prospects for the remortgage product and create scores for each customer....Amongst the results that made a real difference to the business were that they...Achieved with the model, a nine times greater response than that achieved by the control group [and] Secured £33million worth of mortgage application revenue....Clementine has already paid for itself many times over and will continue to do so for many years to come. »

M&T Bank

- « M&T Bank, headquartered in Buffalo, NY is major player in retail banking holding assets of 51.5 Billion as of Q3 2002. When debit cards were first introduced into the market place, research showed that once someone started using a debit card, they didn't stop. M&T wanted to increase debit card use for POS transactions among current customers who had not yet used their card. The key was getting customers to activate the card. Promotions executed through VISA offered a coupon for a free Sara Lee pie as an incentive to start using the card. While the response rate was 14.4%, M&T wanted more. M&T wanted to lower the cost, the turnaround time, and develop their existing customer base. M&T turned to ASA, and the refined analytics of ModelMAX®, for help. Using the data from the Sara Lee promotion, ModelMAX® predictive analysis determined that 55% of all responders could be found by mailing to only 30% of the file. So when ModelMAX® predicted a 27.6% response rate and actual results came in at 25.3%, M&T knew they had something. In fact, these newly activated customers indicated that they didn't know how they got along without the debit card. »

PrimeCredit Limited

- « With more than 1,200 personal financial services companies competing in Hong Kong, PrimeCredit Limited recognized the need to devise and implement an analytical CRM solution that would help the company establish a strong foothold in the local market and gain a competitive edge....As a director of marketing for PrimeCredit in Asia, Daniel Chong realized that the most profitable way to acquire and retain customers would be to make use of an invaluable resource – customer data....”Due to the lack of local public information, it is difficult to go out and acquire demographic data the way companies can do in many other parts of the world,” Chong adds. “Therefore, the data we build up internally takes on an exceptional importance. With the right approach, the correct infrastructure and the best analytical solution, we can use this data as a real resource that can give us a huge competitive advantage.” The newly installed SAS data warehousing system houses the information in a manner that makes it instantly accessible and easy to analyze and monitor. At the same time, SAS Enterprise Miner data mining software helps to reveal trends, explain known patterns and identify critical factors that can secure a desired effect on a company’s bottom line. “There has been obvious improvement in the response rate of PrimeCredit’s direct mailings, because our marketing offers are now much more targeted. We have seen an increase in customer feedback after sending out only one-sixth of the previous number of mailers to selected potential customers,” Chong says....Data mining segments PrimeCredit’s customers into profitability groups to develop cost-effective service levels and targeted marketing campaigns. It also will be used to develop customer scoring and retention models in cross-selling and up-selling strategies....Another advantage of SAS data mining is the ability it gives PrimeCredit to conduct internal analysis on the profitability of its product and branches, allowing the company to identify underperformers and evaluate the ratio of operational costs to revenue generated by them. »

BankFinancial

- « Using SPSS Inc.’s Clementine, BankFinancial is able to get more accurate results in less time, thereby improving the efficiency of its marketing efforts while keeping costs down....”With Clementine, BankFinancial is able to create targeted lists in less time with more accuracy,” said Matt Jacobs, research analyst at BankFinancial. “We now have the ability to make distinctions among our customer base, providing us with a clearer picture of the people we want to reach....Targeting the right people is critical, we have been able to return the same number of responses from a mailing that is only about one-fifth the size,” said Jacobs. The bottom line is more cost-effective direct mail campaigns. By better targeting its mailings, BankFinancial has reduced costs by 79%....The Research Department has seen a dramatic reduction in processing and analytical time. “We’ve cut our time by 80%, which really reduces

our customer acquisition costs,” reported Bill Connerty, assistant vice president at BankFinancial....A major goal is to expand BankFinancial’s base of commercial clients....”We have all kinds of data from our customer satisfaction and product development surveys that we’re going to import into Clementine and come up with a profile we can apply toward our business lists,” said Connerty. “This will allow us to create a business call list we can pass off to our commercial sales group...” BankFinancial is also working to maximize its customer retention efforts. BankFinancial is using Clementine to develop an early warning system that will identify which customers are likely to leave the portfolio and what actions will most likely change their minds....BankFinancial is also working to improve its ability to reach its most valuable customers....BankFinancial is using Clementine to build models that more accurately represent these types of customers. »

Kookmin Bank

- « Ranked the 17th in Asia and the 68th in the world with approximately \$147 billion in assets, the new Kookmin Bank is positioned as a world-class bank. The bank provides a full range of retail and commercial banking services to more than 14 million customers....Teradata Warehouse Miner^[20] solved two important business challenges. First, a significant portion of the bank’s revenue comes from overdraft protection for personal accounts. We used Teradata Warehouse Miner to identify those account holders with good credit ratings who had not yet set up overdraft protection. We approached them about setting up the new services, and this has helped us expand our relationships with existing customers and greatly boosted profitability. Second, to increase sales of personal lines of credit, we use Teradata Warehouse Miner to identify those customers most likely to use this type of loan product. Then, through a rigorous analysis of customer banking behavior and credit ratings, we identified the most appropriate customers and made a loan offer with a reasonable credit limit....Teradata Warehouse Miner provided the data preprocessing and analytical modelling techniques that enabled us to build three models and smoothly integrate them into our business process. »

NASDAQ

- « In this era of Enron, Martha Stewart, and a host of Wall Street scandals, there is a growing need for better tools to monitor the market for suspicious activity that warrants closer inspection. The clues contained in the millions of trades, wire stories, and SEC filings each day make it impossible for humans alone to sift through all the data to perform surveillance....To mine the data, NASD has developed an intelligent surveillance application -- the Securities Observation, News Analysis and Regulation (SONAR) system -- that automatically monitors the NASDAQ, OTC, and futures markets for suspicious patterns. SONAR has been in use since December 2001. Each day it processes between 8,500 and 18,000 news wires stories, approximately 1,000 quarterly and annual SEC filings from corporations, and evaluates price-volume models for 25,000 securities. The system generates 50-to-60 alerts ("breaks") per day for review by several groups of regulatory analysts and investigators. After review, a number of these are referred to the SEC or Justice Department for prosecution. While NASD must be circumspect in describing in detail the efficacy of SONAR, it has released some information: to date, 180 cases have been referred to the SEC or DOJ, including several very high profile cases. AI in the software SONAR includes several AI techniques, such as data mining, natural language processing for text mining, intelligent software agents, rule-based inference, and knowledge-based data representation.. »

Central Institute of Mathematics in Economy, Moscow

- « The goal of this research was to develop an optimal strategy of capital management in this market [Russian Short Term Government Bonds], which would provide the maximum

profit...The data included all bond issues prices, currency rates fluctuations, and other market indices for each trading session, since the establishment of this market up to the present day. These data have been analyzed with the help of PolyAnalyst,^[21] which discovered an algorithm for making a decision of selling or buying various issues of STGB....The PolyAnalyst algorithm efficiency on average exceeds 30%, while for some periods it is close to 60%. An obvious advantage of the present method of the market analysis is the ability of PolyAnalyst to update the algorithm every time there is a need for this. This feature proves to be crucial in cases of sharp perturbations, which are quite common for the Russian financial markets. »

Hang Seng Bank

- « Hang Seng Bank – Hong Kong’s largest locally incorporated bank – has chosen SAS data mining technology to provide the best tailor-made services for the bank’s customers....The information generated from in-depth analysis of customer data – such as product and service combinations, channel usage patterns and responses to marketing campaigns – is required for strategic intelligence. Armed with this intelligence, Hang Seng is able to deliver more personalized services to current customers. Tailoring products to the needs of clients has led to greater customer satisfaction, higher customer retention and increased bank profits. »

Fleet Bank

- « Fleet recently [late 1990’s] used data mining to identify the best prospects for its mutual fund offerings [and indeed all of its retail products]. Grossman’s team mined customer demographics and account data including transaction activity and account balances along several product lines. From that analysis, they found customers who were likely to invest in mutual funds, and they used that information to help Fleet’s Investment Services division target prospective clients. »

JCB Co., Ltd

- « As the largest credit card issuer in Japan, JCB Co., Ltd. has established itself as an international brand. The company, with 34 million cardholders, offers more than 200 services and 600 types of cards. Little by little,... JCB has been testing the waters of data mining technology with projects that include everything from protecting customers’ vital personal data from hackers to identifying potential catalog recipients. In July 1999, JCB’s sales department began a new data mining project aimed at tracking increasingly diverse and complex customer needs. The new system, implemented in only three months with the help of SAS Enterprise Miner, is integrating such cutting-edge scientific marketing concepts as customer profiling and customer relationship management into JCB’s standard business practices. Already the results are impressive. With data mining, JCB has quadrupled the rate of customer responses to direct-mail solicitations and initiated a successful campaign to retain current cardholders –achieving a success rate six to 10 times greater than that of earlier campaigns....Under the new system, JCB uses data mining technology for two purposes. First, it segments customers to increase the response rate of its marketing campaigns, which in turn increases revenue. Then it uses customer profiling technology to devise a customer-focused sales strategy. The system analyzes how members use their cards, helping JCB identify and retain its most profitable customers....Not only are direct-mail campaigns showing much greater success, Enterprise Miner is identifying new segments of potential customers that traditional marketing methods had overlooked. »

AXA Financial Inc.

- « By using SAS’ Enterprise Miner software, AXA’s Customer Segmentation and Analysis Group can determine more quickly and efficiently which customers hold what products and

predict customer preferences for new offerings....the group was able to show that 80 percent of AXA Financial's cross-sells come from the top 30 percent of customers, resulting in considerable savings in cost and time – and increased profitability....The Customer Segmentation and Analysis Group then leverages the sophisticated modelling capabilities of Enterprise Miner to develop cross-sell and up-sell models based on 350 variables per customer – with more than 20 million transactions per year...Satterlee – along with Clarissa Tang, director of marketing tracking and analysis, and Melissa Corso, director of customer modelling and analysis – has brought value to the field offices by providing scored client bases to 5 percent of AXA Financial's professional force in just three months. They can demonstrate which customers are more likely to surrender policies, for example, based on their previous payment history....AXA Financial is also redeveloping relationships with former policyholders. Through modelling, they were able to identify the best prospects and pair them up with financial professionals. As a result, sales in this group have increased by as much as 276 percent. »

Mellon Bank

- « At Pittsburgh's Mellon Bank, IBM's Intelligent Miner is being used...to study as much as 10GB of information on consumer bank customers, with an eye to retaining the most profitable ones. Based on historical data, the bank tries to predict which customers will be profitable in the future. The bank also uses data mining to project which customers are likely to switch from Mellon credit card to another bank's card based on historical patterns of use. Data mining showed that the best predictive factors for Mellon credit-card customer attrition are the frequency of card use and the types of purchases that are or aren't made, says Pete Johnson, vice president of Mellon's advanced technology group...."Data mining helps us play in the national credit-card market more effectively....Small credit-card companies that are unable to embrace this technology can no longer survive because they can't manage their customer bases." »

Chelsea Building Society

- « London 15 September 2003 - Qbase Data Services Ltd today announce the completion of the first stage of a predictive analysis project for the Chelsea Building Society, using KXEN's predictive modelling software. Predictive analytics has been applied to mortgage enquiries, scoring and ranking enquirers based on their propensity to take out an application. Qbase developed a number of models - utilising the KXEN modelling software - which included attributes from internal variables, captured by the Chelsea Building Society's call centre, and external variables from EuroDirect's CAMEO lifestyle classifications. Using KXEN, applications were also scored to predict the likelihood to take out a mortgage. »

Skandia Bank

- « Skandia Bank, which started in 1995, has 177,000 customers. Earlier experience with marketing campaigns showed very thin results....With Skandia Bank's new customized data mining tools, it is possible to focus the sight much more carefully. After 36 campaigns, the Bank can now see the results. Reply levels have increased from one percent to five percent, which represents a 400% increase in effectiveness....With BusinessMiner,^[22] the data mining tool from Business Objects, Skandia Bank know how to get the right customers, and new ones too. »

Bank of Montreal

- « Bank of Montreal...analyzes mortgage customers' transactions in checking, savings and other accounts for insight into who is at risk of defaulting. The bank was surprised to find that some customers who consistently made their mortgage payments late were not necessarily at a

high risk of defaulting. The bank found that a certain type of customer is in the habit of paying bills late but has the wherewithal to fulfil his or her obligations... »

- « Bank of Montreal mines its 1-terabyte “customer knowledge base” to develop profitability profiles of customers based on multiple factors such as the amount of money in particular accounts, demographic information, the number of monthly transactions and their choice of banking channels (teller, ATM or phone). The Toronto-based bank calculates current profitability profiles of households and produces models to predict the profitability of customers over a lifetime. »

Bank of America

- « At Bank of America, customer service representatives equipped with customer profiles gleaned from data mining pitch new products and services that are the most relevant to callers....Data mining [also] holds great promise in assessing the risk of a bank’s entire portfolio of loans, says Jianmin Liu, vice president and project manager in credit risk management for Bank of America’s mortgage division. »
- « Bank of America identified savings of \$4.8 million in two years (a 400 percent return on investment) from use of a credit risk management system provided by SAS Institute consultants and based on statistical and data mining analytics. This model was used to allow Bank of America to find eligible low income and minority customers to ensure B of A’s compliance with the Fair Housing Act. »

The Dreyfus Corp.

- « To manage customer relationships effectively, The Dreyfus Corp. relies extensively on data warehousing and data mining techniques. “Data warehousing is crucial because it basically stores the memory for our enterprise. So it is important to make sure that memory is accurate,” say Dhore [Senior Vice President]. But just as important is the role of data mining. “Data mining is what provides the intelligence for our enterprise. Just having a data warehouse is not enough.” The Dreyfus Corp. mines its customer base to find out more about client preferences.”...Once we know which clusters our customers are in, we are able to offer them better solutions.”...SAS’ data mining software and Dreyfus promotional campaigns go hand in hand. Data mining techniques enable the company to perform a back-end analysis following each campaign to determine its effectiveness....”That way, we can weed out the campaigns that are not profitable and come up with a number of leaner – but more profitable – campaigns in the future,” says Dhore. Strategies like this are paying off. “By using data mining, customer attrition at Dreyfus has decreased by almost 50 percent, while the average customer balance has increased substantially,” explains Dhore. “In fact, this solution has helped us bridge customer satisfaction with bottom-line success.”. »

Marshall & Ilsley Corporation

- « In financial services, the behavior of particular customers purchasing the same product can vary significantly. For example, a customer who opens a \$50,000 home equity line yet never uses it is a less profitable customer (for that product) than a customer that opens a significantly lower line of credit yet frequently uses it. Identifying which customers would exhibit more profitable behavior in response to targeted marketing programs was critical for M&I to run effective, cost-efficient marketing campaigns....Using the Quadstone System, customer behavior predictions were made in two areas: a customer’s “propensity to buy,” and a customer’s “predictive profitability” – an indicator of whether the services or products the customer would likely buy would yield a low or high profit for the bank. The two groups of information were combined. Each customer was then given a “score” based on both criteria – propensity to buy and predictive profitability. With that information in hand, M&I then sent out

a set of three targeted mailings. The first mailing round targeted those high potential profit customers with the product they were most likely to buy. A second and a third round followed, targeting those same customers, if they had not bought the first product offered with the products they were next most likely to buy. And for each new wave of the campaign, depending on how customers responded, M&I analysts were able to change each customer's scores based on their up-to-date predictions and profile and alter their campaigns accordingly....Having a constant set of changing data allows M&I analysts the opportunity to constantly assess customers' needs and tastes and respond more efficiently. »

Stock Selection Using Recon^[23]

- « When trading costs were taken into account, Recon's portfolio [of 50 stocks] had a total return of 238% over a four-year period, significantly outperforming the benchmark [S&P500], which returned 93.5% over the same period. »

Health Care

Pediatrics^[ii]

- « In a data mining study of neonatal outcomes that points to the potential of electronic health records (EHRs) to advance patient care, Pediatrix Medical Group researchers have identified a life-saving difference between two antibiotic combinations frequently used to fight infection in premature infants. Researchers with Pediatrix, the nation's largest provider of neonatal and maternal-fetal care, mined the group practice's clinical database to compare the outcomes of more than 128,000 neonates treated with antibiotics over an eight-year period, 1996-2004. Their peer-reviewed study, "Empiric Use of Ampicillin and Cefotaxime, Compared With Ampicillin and Gentamicin, for Neonates at Risk for Sepsis," appears in January's [2006] Pediatrics journal. In examining outcomes, the Pediatrix study found that premature babies who received the antibiotic combination of ampicillin and cefotaxime within three days of birth had a higher mortality rate, compared with those treated with ampicillin and gentamicin. Based on those results, Pediatrix strongly advises neonatal health professionals to avoid the ampicillin-cefotaxim combination and consider alternative antibiotic treatments. "We are at the early stages of tapping the potential of our data mining model to advance neonatal care on a global basis," said Alan R. Spitzer, M.D., Senior Vice President, Director of the Pediatrix-Obstetrix Center for Research and Education and co-author of the study....Pediatrix launched its antibiotic study following a request from the National Institutes of Health (NIH) and the Food and Drug Administration (FDA) to research the medications most commonly used in neonatal intensive care units (NICUs) based on Pediatrix's outcomes database. That database - the largest in the world - incorporates outcomes from Pediatrix's integrated national practice network, which includes more than 375,000 patient records, representing five million patient days and 1.3 million medication records. "From a broader perspective, the growing use of the electronic health record signals that medicine is now at an exciting threshold," Spitzer added. "We believe that more uniform data collection, made possible by the continuing shift to EHRs, can result in significant new discoveries by medical researchers throughout the world."..."Our database is a renewing source of research and quality improvement programs aimed at reducing the impact of premature birth," said Reese Clark, M.D. Pediatrix's Director for Research and Education. "The value of our database comes in sharing and comparing that data across the organization, which allows for benchmarking on a national level and collaboration with other healthcare providers to improve patient care." Historically, the healthcare industry has relied on extracted, secondary data from financial claims or coding datasets. Pediatrix's database consists of primary data, meaning that the person entering data is also the person providing the care and monitoring the data daily, thereby reducing errors and enhancing clinical care." As a physician-

driven record, the EHR will become the fundamental element in future patient outcome studies for many fields of medicine," said Spitzer. "Consistent patient information is essential for building accurate databases and finding meaningful diagnostic and treatment patterns." Practicing in more than 220 neonatal intensive care units (NICUs) across the country, Pediatrix spent a decade building and maintaining its database, which captures detailed, privacy-protected information about babies' medications, mothers' histories, procedures, and lab and coding information. »

Children's Memorial Research Center

- « Children's Memorial Research Center, a leading U.S. pediatric hospital and research institute, is leveraging SPSS predictive analytics to differentiate pediatric brain tumor types. By analyzing both patient data and medical research literature..., Children's Memorial Research Center has gained unique insights into tumor classification and treatment strategies....Dr. Bremer [director of the Brain Tumor Research Program at Children's Memorial Research Center (the research arm of Children's Memorial Hospital)] uses SPSS data mining technology, Clementine®, to analyze and classify pediatric brain tumor types. Additionally, he employs SPSS text mining software, LexiQuest Mine™, to discover previously overlooked relationships contained in volumes of biomedical literature....Dr. Bremer cited an example from a pilot project where he was able to process 124,000 medical abstracts in exactly one hour and 18 minutes. He contrasted this grid-enabled analysis with the average time it would take him to conduct identical analysis on a stand-alone computer, between 20 and 24 hours. "The automated extraction of information from biomedical literature promises to play an increasingly important role in text-based knowledge discovery processes," reported Dr. Bremer. "Leveraging SPSS predictive analytics and United Devices' expertise in grid computing, we've developed an integrated technology system that can efficiently extract and organize gene relationships from full text articles. We can also correlate this insight with both past and ongoing research on effective pediatric cancer treatments." »

St George's Hospital NHS Trust

- « The shortage of intensive care beds in Britain became a hot political issue in the NHS winter crisis of two years ago when a flu epidemic killed scores of elderly and vulnerable people....The hypothesis was that there would be a marked difference in death rates of intensive care patients they categorised as "at risk" – a third of the total – compared with those who were not....With the help of SPSS software, the researchers created a triage model which took into account the patient's age, disease, length of stay in the unit, whether they had heart/lung surgery and other factors in order to predict who was most in danger....The model was developed from 5,475 patients discharged from Guy's intensive care unit....The model was then applied to 8,449 patients from 20 intensive care units. The post ICU discharge mortality rate of the cohort was 11.3%. 34% of the cohort was identified as at risk. Among these 2875 at risk patients the mortality was 25% (718). Among those not at risk the mortality was only 4%. If the at risk patients were to stay another 48 hours, their mortality could be reduced by 39% (280). To do this would require an increase in the number of fully funded ICU beds by 16%....This is the first study to pin-point a major cause of post ICU deaths in the UK and it would have been difficult, time consuming and expensive without the use of SPSS software. »

Highmark Inc. [liii](#)

- « Pittsburgh, Pa.-based Highmark Inc. uses a data mining tool from Cary, N.C.-based SAS Inc. to analyze billions of bits of information about its patient population. The software alerted the payer that there was a link between glaucoma and coronary heart diseases and the correlation was stronger under certain circumstances, explains Richard Pro, vice president of health care informatics research and analysis. "We were not looking for such a connection up front when

we analyzed the data-the software found the link and showed us where the correlation was strongest," Pro says. "Data mining is not like querying a database-a query is a question built on an assumption. Data mining software actually tells you something about your data that you didn't know or ask."... At Highmark, data mining efforts have enabled the payer to not only identify which patients need interventions and case management, but which patients need it the most, Pro says. "Data mining feeds our predictive modeling efforts, but it's not enough to just say that 'these patients likely will have problems in the future'-we have to pinpoint what types of problems they will have to discover what is causing them," he says. For example, a predictive modeling application might give a high score to a diabetic patient whose claims history indicates that they do not have their disease under control and will be costly to treat in the future. But Highmark's data mining has found that another group that is at risk for contracting diabetes-patients with certain cardiac conditions who are on drug therapies that substantially increase the chances of them developing diabetes down the road. "One group would obviously be monitored for diabetes, but if we hadn't made that connection for the other group, diabetes wouldn't be a focus of their treatment-we would be managing one morbidity without paying attention to another," Pro explains. "Many complexities of patients' conditions have been documented by medical research, but they have not been a part of the business side of health care. We haven't had the capability to use software to connect all the dots and understand when and how we can intervene. If you're going to go out and tell patients they must be in a care management program, you have to know your predictions that they are in danger of health problems are accurate. Data mining is the foundation for predictive modeling because it locates and confirms those connections."... Most payer organizations use third-party fraud detection software that analyzes data using an engine of known fraud scenarios and see if claims submissions fit those patterns. The problem is that once fraud patterns are known-and software is designed to detect them-the horse has already left the barn, Pro says. "By the time fraud detection software can identify a pattern of fraud, it's already been going on for quite a while and you've already lost money as a result," he says. "Plus, new types of fraudulent activities are cropping up all the time, and a managed care organization can't really wait to get a call on a hotline to do something about it." Highmark has used its data mining software to develop a neural network, a form of artificial intelligence, that monitors claims submissions and looks for known patterns of fraud, as well as subtle hints of suspicious claims activity. The network channels flagged claims through filters that pare down claims data to focus on what types of information or activity are deemed suspicious. That filtered data then is passed on to a team of investigators at Highmark that examines specific elements of the claim. Pro declines to provide details on what types of fraud the data mining software has unearthed. But he says the payer's efforts have enabled it to identify unique patterns of fraud that had not yet been flagged by commercial fraud products. "Data mining is really starting to pay off for us when it comes to fraud detection," he adds. "The power of the technology is that it doesn't look for established patterns-it goes out and finds them." . »

Pfizer, Inc.

- « Pfizer, Inc., a major research based, global health care company, is at the forefront of research on therapies for male erectile dysfunction (ED). Recently, the company has received FDA approval for a new treatment called Viagra, the first oral treatment for this condition....Pfizer awarded an ED research grant to a team led by Dr. Raymond C. Rosen, an internationally recognized ED expert. The research resulted in the development of...the International Index of Erectile Function (IIEF)...this questionnaire is a self-administered, 15-item measure that is cross culturally valid and psychometrically sound...Pfizer led a worldwide market research effort...to determine the IIEF's usability... The overall findings indicated that an abbreviated version of the IIEF would further increase acceptance by doctors and patients...Pfizer then tasked its researchers to use proven statistical methods to

reduce the 15-item IIEF to five questions that would conform to the National Institutes of Health (NIH) definition of ED...The SHIM:IIEF-5 was developed using data from four major studies of men diagnosed with ED and two control samples of men without a history of ED....The data were analysed using CART^[24] and logistic regression methodologies in concert. CART was used to rate the relative importance of each of the IIEF's 15 items in terms of their ability to discriminate between the presence and absence of ED....Dr. Cappelleri and his Pfizer colleagues found firm agreement between the CART results and the NIH definition of ED....The next step after selecting the questions was to develop a scoring system that would be easy to administer. In this case, Dr. Cappelleri wanted to determine a cut-off point in which men scoring at that point or lower on the SHIM:IIEF-5 could be classified as having ED, while men scoring higher could be classified as having normal erectile functionality....Dr. Cappelleri then used CART to develop a scoring system to determine an objective SHIM:IIEF-5 score that gave a high level of sensitivity (high probability of correctly identifying ED) and specificity (high probability of correctly identifying men without ED).... As a tool in the identification of such an under diagnosed condition, the SHIM:IIEF-5 is a crucial part of the Outcomes Research Tools and marketing programs for Viagra. »

Olympic Games 2002 and State of Pennsylvania

- « The U.S. has never seen security measures like those now in Salt Lake City for the Olympics. One little known measure, just put into place, gives early warning of a possible bioterrorist attack. The Artificial Intelligence (AI) computer system analyzes patient data from emergency rooms and instant care facilities across the state. If it detects a significant pattern, it pages the on-call state public health physician...It takes data from 80% of the health care systems in the region, including Intermountain Health Care, the official provider of health care for the Olympics....RODS [Real-time Outbreak Detection System] is an early warning system to assist health care personnel in the early interception and treatment of an outbreak – not just from bioterrorism....The RODS system collects and analyzes relevant data such as emergency room registration data and lab results. It uses several Artificial Intelligence techniques for machine learning, natural language and data mining....RODS was initially installed in Western Pennsylvania in August of 1999 for public health surveillance for the 3 million residents of 13 counties. So far, the system has detected only a naturally occurring outbreak of influenza. It is now being extended to all of Pennsylvania. Other public health departments in the U.S. are now showing interest in RODS. »

Universidad de Santiago de Compostela

- « CleverPath Predictive Analysis Server has delivered very promising results,” said Juan José Casares, the chemical engineering professor who is directing the research. “By enabling our researchers to predict high-ozone events, we can protect the residents of our region from potentially dangerous atmospheric conditions.”...By tracking the factors that influence ozone – sun radiation, air temperature, relative humidity, wind direction, and the concentration of nitrogen and volatile organic compounds – US researchers believed they could build an increasingly accurate model to predict ozone levels. The USC team selected CA's CleverPath Predictive Analysis solution over several alternatives for its ability to combine predictive capabilities with real-time presentation of information. Researchers were able to feed historical ozone readings into the self-learning capabilities of CleverPath Predictive Analysis Server – rather than having to craft the mathematical model manually. The ability to learn from past data allowed the team to achieve effective prediction capabilities in record time. Within four weeks the CleverPath-based system was able to predict ozone levels on a quantitative basis for three to six hours in advance, and was able to offer trends over longer periods of time...The initial [result] of this project has encouraged the University to extend the same technique to other

polluting agents, such as sulfur oxides, nitrogen oxides, and pollen. »

American Healthways

- « To augment the company's ability to prospectively identify high-risk patients, American Healthways uses predictive modelling technology from SAS. "Essentially, predictive modelling helps us identify patients who are trending toward a high-risk condition, which gives our nurse care coordinators a head start in identifying high-risk individuals. Then we can take steps to improve their quality of healthcare and prevent health problems in the future," explains Mark Ridinger, MD, the company's senior vice president and chief science officer..."Everything we do is ultimately designed to facilitate the patient-doctor relationship," says Ridinger. "By melding high tech solutions like predictive modelling with our clinical information system in support of our nurses, we're working with these key stakeholders to close the gaps that exist in the healthcare system today..."the predictive model developed with SAS generates risk scores for each patient. American Healthways nurses then initiate patient relationships, armed with better knowledge about the likely course of the patient's disease. Once a relationship is established, the company collects additional information from these patients, including family history, eating habits and current health concerns..."Our intervention piece is what really distinguishes us from our competitors," says Ridinger. "Other companies may use predictive modelling, but they don't interact with the patients. And you're never going to make a difference if you don't intervene. On the flip side, some companies have intervention programs but aren't maximizing the value of prospectively identifying high-risk individuals....Enterprise Miner not only provides an ability to do artificial intelligence and neural network, but to do many other important statistical procedures as well....What we get done in three days with SAS would have taken weeks to do before....The efficiency of setting up the process and comparing the models is pretty dramatic." »

New York City Health Department

- « After Sept. 11 and the anthrax scare, health officials nationwide began clamoring for a better data infrastructure so they could view the same health data that the first responders in emergency rooms and ambulances were seeing....They are struggling to develop an information infrastructure that will link emergency rooms, physicians and public health departments, and serve as an early warning system for public health crises and bioterrorism....One such early warning system is already in use by the New York City Health Department. Epidemiologist Dr. Farzad Mostashari oversees a citywide health surveillance network that links emergency rooms, 911 dispatch facilities and pharmacies. New York's system can detect phenomena such as widespread outbreak of non-specific symptoms – coughs, sore throats and fevers, for example. A Microsoft SQL Server database at the Health department stores ambulance dispatch data. It arrives automatically by File Transfer Protocol as delimited text files over CityNet, New York's data communications backbone....The database receives about 3,000 records a day. Health Department officials study the records using a data analysis program from SAS...Next, a predictive model compares the number of dispatch calls received in a 24-hour period with the number expected based on call histories...."We're moving towards the same kind of system with city employee sick logs and also pharmacy sales of antidiarrheal products," Mostashari said. "If Kaopectate starts flying off the shelf, that could be cause for alarm..."A cluster of emergency room visits and ambulance calls for respiratory distress did occur in Queens one day after the crash of American Airlines Flight 587 on Nov. 12...The cause turned out to be smoke inhalation from the crash and not bioterrorist attack. But it proved the system could alert officials to a time-space cluster of health anomalies. »

First Choice Medical Management

- « First Choice Medical Management (FCMM) is a leading-edge provider of healthcare billing

and practice management services in Texas...FCMM staff use PolyVista's OLAP functionality to perform ad hoc queries/reports or to analyze or validate certain types of transactions....FCMM staff use PolyVista's Discovery Algorithms^[25] to proactively uncover data anomalies, which can be indicative of unfavorable trends, data-entry errors or payer discrepancies. Without identification and correction, these errors can significantly delay payments and will cause additional follow-up work. FCMM staff use OLAP and Discovery to better manage and analyze their billing and collection function for productivity, accuracy and quality, and to monitor billing data for regulatory compliance....Physicians, business managers and administrative staff use PolyVista to enhance productivity and optimise professional staffing requirements, determine procedural and fee-related trends and patterns, and to analyze patient billing histories and insurance claim response. »

Baylor Health Care System^[iv]

- « By using SAS Enterprise Miner in conjunction with other integrated components of the SAS9 Intelligence Platform, researchers can measure and analyze quality improvement efforts related to the provider's \$119 million clinical transformation initiative. In 2002, Baylor Health Care System began a seven-year effort to redesign its clinical processes and to seamlessly link information throughout its network of 11 hospitals, 69 clinic sites and numerous physician practices. This clinical transformation includes investments in IT and changes in the way patients are treated - all with an eye toward using evidence-based decision making to improve care. The goal of the clinical transformation is to generate a single, secure electronic health record (EHR) for each patient treated in the Baylor Health Care System. Accessible through a handheld or desktop computer, these EHRs will be extremely valuable to clinical care providers by improving the collaboration between clinicians in the management and delivery of patient care. The use of EHRs will also produce a deluge of clinical data that can be used by biostatisticians and clinical scholars using SAS software to conduct research pertaining to healthcare quality and effectiveness. The insights derived from this research will be deployed in the operationally focused healthcare improvement initiatives of Baylor Health Care System. Baylor Health Care System will also use SAS software to sift through vast amounts of data generated internally and available from external sources, including the Medical Expenditure Panel Survey (MEPS), Texas mortality data and national Medicare claims data. By analyzing information from a wide range of sources, clinical scholars will make an effort to identify certain segments within Baylor Health Care System's patient population that may respond favorably to clinical intervention. For instance, analysts will use SAS in a randomized trial to evaluate a new program that provides transitional care to elderly patients moving from in-patient to discharge status in an attempt to determine if these treatments can improve outcomes. »

University of Sheffield

- « Analyzing and interpreting the huge quantities of available data is a taxing problem for today's diagnostician. Using traditional spreadsheets and tables, it is difficult to identify the trends, anomalies, and other significant features in the data that are essential to accurate diagnosis. The University of Sheffield team, led by Dr. Rob Harrison of the Department of Automatic Control and Systems Engineering and Dr. Simon Cross of the Department of Pathology, set out to create a more effective diagnostic tool....Neural networks were ideal for this type of work, explains Harrison: "They can cope with noisy data and, unlike some other systems, they begin by assuming nothing, so they can sometimes identify anomalies that might defeat other systems. They provide a practical, day-to-day tool for reaching the conditional mean without becoming bogged down in theory."...Standard neural networks deal with multi-dimensional data, which is difficult for the human brain to assimilate. MATLAB^[26] reduces

this data to 3-D graphs and plots with overlay that can be readily understood even by inexperienced observers....The technique they developed uses a growing cell structure (GCS) network, an unsupervised self-learning artificial neural network....To test the technique, the researchers selected 692 breast lump specimens, noting the patient's age and ten other pertinent variables. They used either an open biopsy or information from mammograms, combined with the absence of further malignancy, to confirm the final outcome of benign or malignant cancer....MATLAB produced a network that contained areas of high certainty positive and negative outcomes, with a narrow area of transition in between. This result compared very favorably with that obtained using the conventional logistic regression technique (96% as opposed to 98%). "The new technique does not replace human interpretation of data, but it provides a powerful decision support mechanism to help validate the results," says Dr Cross. "It is the most useful technique we have had for years." »

US Defense Department

- « Dr. Scott Optenberg and associates with the Civilian Health and Medical Program of the Uniformed Services (CHAMPUS) and the Department of Defense (DoD) are using SAS solutions to spot billing and health services fraud....By utilizing data warehousing and mining tools built at their Fort Sam Houston, Texas, facility, Dr. Optenberg and his colleagues brought an end to a crime spree for an offender who was double-billing CHAMPUS. At the same time the provider was submitting bills to CHAMPUS, he was seeking payment for the same services from the Medicare program. The fraudulent claims submitted from this single vendor totaled \$1.7 million. »

National Research Center for Surgery, Moscow

- « A number of clinics around the world perform heart transplantation. One of the most difficult problems in this field is rejection of the transplanted tissue. It is critically important to know the degree of this rejection. Today the most accurate method to measure it is biopsy...However, making a biopsy of the heart tissue is a complex, dangerous, and undesirable action. In the present study, each time the biopsy was made, the blood sample was also taken. Each blood test provides a great number of immunological parameters, which may reflect the degree of rejection. Taking blood samples is obviously much easier than performing a biopsy. Thus, the possibility of predicting the rejection from immunological data would be extremely beneficial for patients with a transplanted heart...The results obtained by PolyAnalyst have a surpassing scientific value and large clinical implications....PolyAnalyst obtained a nonlinear single-parameter formula which provides correct prediction of the rejection degree with an accuracy of about 70%. This model has been clinically used in NRCS for more than a year, allowing to reduce the total number of heart biopsies by 44%. »

University Clinic for Anesthesiology, Universität Ulm

- « Critically ill patients in intensive care often require artificial lung ventilation....Different patients require different regimes of the artificial lung ventilation. It is very important to perform classification of patients into groups requiring similar regimes of ventilation. PolyAnalyst processed the data from the computerized system "RICA", developed at Ulm university, that contained data from monitoring devices, blood analyses, results of doctor examinations, XR, tomography, etc....PolyAnalyst has developed a set of nontrivial rules for each of classification tasks. These rules allow correct diagnostics in the 94% of cases....We hope to be able to make further use of assistance...in building a fully automatic system to assess the respiration of an intensive care patient. »

San Francisco Heart Institute

- « The institute also does physician profiling to review variations in performance, assessing how

length of stay, outcomes and cost vary from physician to physician...By analyzing the factors involved, the hospital shortened the average stay from seven to three nights. »

Anthem, Inc.

- « But mining the combined data in its 1.3 terabyte NCR Teradata database helped the \$6 billion health insurance company target areas to trim – such as coronary artery bypass, a common but expensive and risky procedure. Eighteen different hospitals within a region of Ohio were performing bypasses but the variation in outcomes among them was striking. Anthem chose five hospitals as centers of excellence and shifted all of the bypass patients in the region over to them. As a result, costs fell dramatically and the mortality rate was reduced from more than 4 percent to less than 1 percent, says Anthem's senior vice president, William R. Milnes Jr. »

Bridgeport Hospital

- « An estimated 5 million Americans rush to emergency rooms each year with chest pain. The \$64,000 question for physicians is: Which of these heart attack candidates should be admitted and who can safely be sent home? “To date, there has been a lack of definitive diagnostic tests for early detection of acute myocardial infarction (AMI),” says Dr. Larry Bernstein....Consequently, most people even suspected of having a heart attack are admitted to CCU for a two-day observation to rule out AMI, averaging up to \$15,900 per patient. About 68 percent of these stays “rule out” patients who were not experiencing AMI and could have received more cost-effective care. Conversely, about 34,000 patients are released from hospitals with undiagnosed AMIs, which can prove deadly. The result is that each year in the U.S. alone cardiovascular disease carries a direct cost of \$48 billion and additional indirect costs of \$54 billion....Bernstein is one of a handful of researchers who think the answer to this problem may have finally arrived. Just last month (November 1995), Beohringer Mannheim received FDA marketing clearance for its proprietary method for detecting troponin T, CARDIAC T. This easy-to-use diagnostic test is based on the company's marker, cardiac troponin T....Bernstein's goal was to classify patients according to their symptoms upon admittance, and test results for two traditional AMI diagnostic tests (CK and CK-MB) and troponin T....Bernstein accomplished this by looking at the medical records of patients admitted as heart attack and performing an outcomes analysis based on dividing the patients into two groups: heart attacks and not. “S-PLUS^[27] made it easy to set up the matrix of patient data, then analyze the ranges of all three tests on both groups using continuous variables to determine what a normal or high response to each test is...Bernstein's study concluded that using a combination of the Goldman algorithm and troponin T tests in a systematized procedure may be a viable possibility of improving the treatment of patients with chest pain and reducing over utilization of CCU....Bernstein and his colleagues also examined the fiscal year 1994 Bridgeport Hospital report of length of stay (LOS) and charges....They estimated that the hospital could have eliminated unnecessary hospital stays for chest pain patients and reduced CCU stays for possible angina patients resulting in a combined reduction in charges totaling \$360,563....”This test is as easy to administer as a pregnancy test, and could significantly reduce the costs related to unnecessary CCU stays,” says Bernstein. »

Olsten Health Services

- « In the health-care industry, however, running out of a critical medicine at the wrong time can mean the difference between sickness and health...Olsten Health Services mines its data to predict what products and services it will need in the future based on utilization patterns from the past. Knowing, say, that hemoglobin levels of diabetics are higher in August than December tells Olsten to stock an adequate supply of drugs and personnel in the summer. »

Cardiff Public Health Laboratory (UK)

- « There is an ongoing infection problem with *Klebsiella* at the University Hospital of Wales... there have been sporadic clusters of colonisation with a few cases of infection from 1995 to 1999. The strains involved were mostly identified to the species *Klebsiella aerogenes* and showed resistance to multiple antibiotics. The data downloaded as input for development of the cross-infection detection program included one of these clusters. This was not actually called as an outbreak, because small numbers of patients were involved, and the organisms were identified as multi-resistant *Klebsiella oxytoca*, rather than *Klebsiella aerogenes*. However, in retrospect, these organisms had closely similar antibiograms and biochemical patterns, and probably represented a cluster of nosocomial colonisation/infection. This cluster, which had not been recognised as such in the laboratory at the time, was strikingly obvious in the teaching set output from the detection program. »

MEDai

- « MEDai, located in Orlando, Florida, uses a combination of artificial intelligence and statistical analysis to provide solutions for the healthcare industry....MEDai also uses its artificial intelligence technology to predict the movement of stocks...MEDai was looking for a way to predict the stock market using chart pattern recognition indicators....the patterns were very subtle and the returns of stocks had such a high standard deviation. Using WizWhy, we were able to find these patterns and generate a significant return on our stock portfolio. »

Public Transport (Air, Rail, etc)

Warwickshire County Council

- « Warwickshire County Council provides public transport for 5 districts from Stratford-upon-Avon to North Warwickshire covering an area of 200,000 hectares and a potential customer base of ½ million residents....They needed to identify rural transport problems, rural isolation and social exclusion issues which residents have encountered whilst using or needing to use public transport facilities....Project officers were consulted to ensure that the questions would give the information needed to take their action plan forward, as well as taking into account the coding required to put it into SPSS effectively....There were over 3,000 responses giving a response rate ranging from 11% to 21% for these areas, with a 17% response overall. Using SPSS to analyse the results significantly found that hospitals were identified as being the most difficult services to access. However, Social Services, leisure/sport facilities, shopping and careers gave particular problems depending on the respondents age or whether they had a disability. By actually discovering these problems meant that an action plan to rectify them was easy to put together and aid in the securing of funding....A consultant would have cost at least £15,000 and the results may not have been available for secondary analysis. The cost for us has been approximately £4,000 in questionnaire design, input of data and analysis using SPSS software. »

Swissair MZ-VPP (Airport Policies and Infrastructure)

- « An important facet of the delay situation at an airport is capacity, which is partly controlled by the acceptance rate (the number of maximal allowed arrivals, prescribed by the tower). In an exploratory data analysis, we investigate the correlation between acceptance rate, schedule, and delays, mainly using visualization methods. Hourly aggregated observations show that the occurrence of delays is clearly enlarged when the number of scheduled flights is close to full capacity. However, this feature is not visible on daily aggregated data. That means that at least an hourly scale is needed to monitor the situation. The axSDatamining team is currently connecting delay reasons and flight schedules. The goal is to obtain a set of elastic criteria

against which to test future schedules. »

Qantas Airways

- « We're currently using KDW [NCR's Knowledge Discovery Workbench] to investigate flight delays and build models to predict whether or not a flight will be delayed at departure and for how long. This can help us in revising our schedules to ensure flights leave on time. »

Southwest Airlines

- « MEGAPUTER Intelligence Inc. announced a successful completion of the proof-of-concept project demonstrating benefits of applying the data and text mining system PolyAnalyst for airline flight safety data analysis. The project was sponsored by the Federal Aviation Administration, Office of System Safety, and carried out in cooperation with the Analytical Methods and Tools Working Group of the Global Aviation Information Network (GAIN)The purpose of the project was to develop new text mining methodologies for the analysis of flight safety data. PolyAnalyst from MEGAPUTER Intelligence was used for the analysis of de-identified safety data from the Aviation Safety Action Program (ASAP) database of Southwest Airlines, with the ultimate goal of further enhancing long-term flight safety performance. "PolyAnalyst enables flight safety officers to quickly perform extraction of stable patterns of terms occurring in pilot narratives and present the results in a convenient visual form," says Tim Logan, Director, Flight Operational Safety, Southwest Airlines. "The visualization capabilities of PolyAnalyst optimize the usage of flight safety officers' time and help create easy-to-understand executive reports, thus making the results of the analysis available to more decision makers in a timely manner."..."The project helped to demonstrate how much potential text mining has to become a key safety data analysis technique," says Christopher A. Hart, Assistant Administrator for System Safety, Federal Aviation Administration. "Exciting advances in text and data mining tools are enhancing the aviation community's ability to convert ever-growing volumes of data into valuable knowledge to further improve an already-commendable safety record." »

Atraxis AG, Swissair Group

- « The marketing department was in an initial phase to build-up a CRM environment, which allows to serve airline customers individually. Thus, the main objective was to enrich the knowledge about individual customers leading to new strategic customer segments. These segmentation results are mainly needed for marketing concerns (e.g. promotions, targeted campaigns....The target population of our project contains not all customers, but the members of the frequent flyer program of an airline....We draw a random sample out of all members of the frequent flyer program database (n=70'000). For these members, we extract all flight activities (about 700'000) as well as all supplier activities (about 300'000) for a 24-month period. We restrict the database to members of certain program countries and members which had at least one flight activity in the period given (about 60% of all members in the sample). The final sample size is about 40'000 customers....We apply a clustering algorithm in order to identify groups which are different from each other according to their product mix as well as to their value, but whose members are very similar to each other....The input variables of the final models are number of trips and segments, the number/proportion of return trips and the number/proportion of within week, weekend, and long stay return trips....The segmentation leads to sensible clusters concerning the business problem....The segmentation are very useful for marketing concerns and for improving customer services. »

IAURIF

- « IAURIF, a French regional studies organization, needed to predict what mode of transportation Parisians would use - and why they would use it - from a large data set not

originally collected for data mining. With Clementine's rule-induction algorithms, IAURIF uncovered unexpected insights, and proved the group's first assumption, which was based solely on experience, to be untrue. Instead, Clementine's rapid modeling environment revealed the most important travel factors, and derived accurate results based on fact. »

Metropolitan Transportation Authority of New York

- « The New York State Metropolitan Transportation Authority is one of the largest governmental organizations in the country. With nearly 60,000 employees and a capital budget of \$12 billion, it is target for internal white-collar crime!...A scheme, which WizRule uncovered, was orchestrated by a supervisor of a large department. Since weekly salaries are calculated to the fraction of a penny, but paychecks are issued rounded to the nearest penny, the MTA issues paychecks on a regular basis which add the accumulated penny earned by the employee to his regular salary. Employees may not notice that the penny was or was not added. A department head with a thousand employees used his computer knowledge to divert the accumulated pennies to his own account, adding several thousand dollars a year to his own income. When run against the payroll database, WizRule found the employees who had not received their pennies, since these were deviations from the rule. In addition, WizRule exposed that the employees were all under the control of one supervisor. This information triggered an investigation. Result: The supervisor was found and prosecuted. »

Pharmaceutical Industry and Bioinformatics

Unilever

- « UK based Unilever's Environmental Safety Laboratory – a state-of-the-art toxicology facility – approves the safety of numerous new and proposed products each year....With Clementine's rapid modelling environment for data mining, Unilever modelled the corrosivity of organic acids, bases and phenols, critical ingredients in many new products. Then putting Clementine's neural network models to work, Unilever trained the models to judge corrosivity based on several descriptive attributes. Clementine's models enabled Unilever to go beyond the limited "corrosive" or "non-corrosive" categories previously established. Since the most strongly corrosive or non-corrosive substances quickly gravitated to their respective extremes, and the substances in between were scored to reveal a gradation of corrosivity, Unilever was able to create a more complete way to test for corrosive substances in new products....With Clementine and this new process, Unilever is leading the way toward *in computero* research, and away from *in vivo* and *in vitro* experimentation. The end result? They saved significant time and money in product development cycles and minimised the need for animal testing. »

Children's Memorial Hospital[\[28\]](#)

- « Each year, nearly 3,000 children in the U.S. are diagnosed with brain tumors. Almost half will die within five years, making it the most fatal cancer among children. If a child does survive a brain tumor, the long-term effects can be significant, and can include neurological disabilities, retardation and psychological problems. Beyond surgery, successful treatments for pediatric brain tumors are rare. Dr. Eric Bremer, director of brain tumor research at Children's Memorial Hospital in Chicago, is one of the leading scientists searching for a better way to treat pediatric brain tumors. One of Dr. Bremer's main goals is to build a gene expression database for pediatric brain tumors, and to then correlate this with both past and ongoing research on effective treatments. As a result of the mapping of the human genome, researchers have gained new tools to study these genetic variations, but the work can quickly produce an overwhelming amount of data. His challenge is to make sense of the 7,000 to

30,000 data points for each brain tumor sample. To do this, Dr. Bremer uses Clementine®, a data mining workbench from SPSS Inc., which enables him to quickly analyze this voluminous amount of data in different ways, and identify patterns and relationships. As an example of Clementine's use, Dr. Bremer combined his own data with that of a publicly available data set resulting in a total of 133 tumor samples from the six major pediatric brain tumor types. Clementine classified these tumors with greater than 95 percent accuracy. He then uses SPSS' LexiQuest™ Mine, a text mining technology, to sift through mountains of scientific literature to extract patterns that, for example, when combined with genetic patterns identified from his brain tumor database with the help of Clementine, can be used to help him evaluate prime drug targets that would form the basis for a cancer cure – Dr. Bremer's ultimate goal....As a result of Dr. Bremer's and other researchers work, types of pediatric brain tumors can be more accurately diagnosed, and the life expectancy for children with brain tumors has grown from five months to 39 months. »

Critical Outcome Technologies

- « The ability to predict biological activity based on molecular structure is leading researchers to breakthroughs in the most complex challenges of medicine. Using a combination of artificial intelligence tools, Dr. Wayne Danter of Critical Outcome Technologies (London, Ontario, Canada) has developed a method to predict whether specific molecular structures are effective against a disease. Currently under study is the HIV1 virus....Modeling each molecule and predicting its effectiveness using standard statistical methods is virtually impossible because of the enormous number of variables. Dr. Danter uses CART (Classification and Regression Trees), a software package from Salford Systems to help build models that isolate the most important variables. Working with public domain, molecular HIV data, Danter trains CART and complementary systems to predict if a given molecular structure is biologically active against a disease....To satisfy Dr. Danter's specialized modelling needs in his HIV research, he inputs the results into another Salford Systems product, MARS (Multivariate Adaptive Regression Splines), then into a neural network program from Ward Systems Group, NeuroShell Classifier.^[29] MARS is a non-parametric regression procedure that extends Dr. Danter's work by improving the accuracy of predictions. NeuroShell Classifier then categorizes a molecule's activity based on patterns derived from CART and MARS....The ability to analyze molecular structure and predict effectiveness helps Dr. Danter look for existing drugs to battle diseases like HIV, as well as to develop potential new medications....The results to date are impressive. In a recent study conducted by Dr. Danter, he analysed 311 drugs with known *in vitro* activity against the HIV1 virus. The system correctly classified more than 96% of the molecules....During the past several months, Dr. Danter has also used CART in developing models to study central nervous system receptors, anti-arthritic medications, and antibiotics, among others. »

AnVil and HealthSouth

- « ...AnVil, a small bioinformatics firm in Burlington, is hoping to unveil potentially lucrative secrets hidden inside the reams of patient data collected by one of the nation's largest network of health-care providers. The company is set to announce today [29 July 2002] that it has reached a deal with Birmingham, Ala.-based HealthSouth to apply its experience in analyzing complex databases for drug researchers to HealthSouth's mammoth repository of patient records. The companies are betting on the analysis, which will focus on drugs to treat stroke and orthopedic patients, to produce insights that will not only help HealthSouth improve medical care but also aid drug makers in their efforts to discover and develop new drugs.... The hope is that such data could reveal how drugs work in real-world settings,

shedding light on questions that are not answered in tightly controlled clinical trials. Researchers could test hypotheses in large numbers of patients over extended periods of time. They could spot subsets of patients who don't respond to existing medications, revealing a promising avenue of research for new drugs or for improved versions of existing drugs.... HealthSouth runs a nationwide network of outpatient surgery centers, diagnostic imaging clinics, and rehabilitative services, treating between 1 million and 2 million patients a year. Given that amount of data, the potential exists to answer questions about the safety and effectiveness of drugs in a way that can not be replicated in clinical trials. »

UCB Pharma

- « The UCB Group is a major Belgian corporation with core business centered on chemicals, films, and pharmaceuticals...its UCB Pharma division researches, produces, and markets medical products covering the central nervous system, the cardiovascular system, and immuno-allerology. The pharmaceutical market has changed dramatically over recent decades....The presence of fierce competition on all sides has prompted UCB Pharma to examine the market very carefully before it introduces any new drug....”Omega, our data mining tool, was developed in SAS software by SPS, a Quality Partner of SAS Institute.”...Omega examines, interprets, and reports on trends in purchasing behavior....[It] facilitates strategic decisions on our existing product range, as well as assisting with new product development and the way we do business generally. »

Human Genome Project

- « The HGP has been trying for more than a decade to decipher the complete sequence of the human genetic code....the hope is that the research eventually will reveal what genetic patterns cause diseases, with the possibility of cures to follow....Bruce Weir, Ph.D., a DNA analysis expert at North Carolina State University in Raleigh, used SAS Enterprise Miner to analyze SNP [single nucleotide polymorphism] data from patients with Alzheimer’s disease. His team, which includes geneticists and statisticians from both the public and private sectors, found genetic patterns associated with the disease. »

NIEHS – Predictive Toxicology Challenge

- « A second round of the challenge (PTE2) [1999] consists of predicting the outcome for 30 chemical bioassays for carcinogenesis...The data provided includes both structural [atoms and bonds making up the molecules] and non-structural [short term toxicity assays] information. ...Provided a sufficiently expressive representation language, good solutions may be obtained from the structural information only. ...Experiments...with STEPS support this claim....The rules obtained by STEPS using structural information only, are comparable in terms of accuracy to those obtained using both structural and non-structural information by all PTE2 participants. In addition, this approach may produce insights into the underlying chemistry of carcinogenicity, one of the principal aims of the PTE2 challenge. Furthermore...carcinogenic activity for a new chemical can be predicted without the need to obtain the non-structural information from laboratory bioassays. Hence, the results may be expected in a more economical and timely fashion, while also reducing reliance on the use of laboratory animals. »
- The Predictive Toxicology Challenge was devised to provide Machine Learning programs with the opportunity to participate in an enterprise of immense humanitarian and scientific value. Details of the Predictive Toxicology Challenge 2000-2001 are at <http://www.informatik.uni-freiburg.de/~ml/ptc/>.

Manufacturing and Production

Hewlett Packard

- « Hewlett Packard uses KnowledgeSEEKER in their U.S. manufacturing plants as a process control tool. One example of the product in action was trouble shooting quality assurance problems in the production process for the HP IIc color scanner. According to Quality Assurance Manager Ray Cole, KnowledgeSEEKER quickly identified the critical factors – several of which had been completely unsuspected – and allowed the engineers to correct the problem. In another case, KnowledgeSEEKER derived the rules necessary to identify situations where a manufacturing process was about to go out of control. This allowed intervening actions to be taken which prevented the problems from ever arising again. »
- « The new HP is the result of a May 3, 2002, merger with Compaq Computer Corporation, with combined revenue of approximately \$81.7 billion in fiscal 2001 and operations in more than 160 countries....There are two fundamental areas in which SAS Text Miner^[30] has helped solve business problems relating to CRM applications at HP. First, our inside call and contact centers have telesales representatives who routinely enter notes into a Siebel application while conversing with customers or prospects. These notes, which contain large volumes of textual information, are usually free-form text and sometimes contain portions of e-mail messages. While it is very useful for the telesales representatives to review these notes in ongoing CRM communications, our internal reporting teams could not come up with an effective report or OLAP application to analyze these notes in collective form. SAS Text Miner was used to effectively partition notes around common themes, which could then be used for subsequent analysis. A second application was one in which the combination of product hierarchies from pre-merger Digital, Compaq and Tandem caused difficulties in analysing customer purchases in like product groups. Because of differing previous hierarchies among the companies, it was difficult to find all of the products in past invoice histories. SAS Text Miner was used to combine the product descriptions at various hierarchies and then create a classification model to determine which product part number belonged to a logical product group such as desktops, high-end servers, portables and the like. There were 25 product classifications, and the classification of more than one million part numbers (SKUs) was completed with an accuracy of slightly more than 90 percent....We are currently investigating the use of SAS Text Miner for analysing notes of product warranty information, which is also free-form text and in multiple languages. »

DaimlerChrysler

- « Automobile manufacturing is inherently a complex process. DaimlerChrysler builds many models of cars, trucks and minivans, each with distinct characteristics. As these vehicles are used in a variety of environments huge amounts of data are produced in terms of warranty claims. The analysis of this very large and complex data set is practically impossible without automated discovery....Warranty claim data is enriched with data from a wide variety of internal and external sources and Information Discovery's system mines and analyses this data periodically, discovering patterns. Every month, this mountain of data is incrementally mined...Refined patterns are combined and stored in the *Quality & Reliability Pattern Warehouse*...Users access the refined patterns, not the data....The Pattern Warehouse automatically identifies the unusual data densities that may be caused by usage or external conditions or those that are tell-tale signs of internal process variations....As new vehicle models are introduced, the Information Discovery *Pattern Warehouse* continuously improves product quality by automatically identifying the key factors that give rise [to] claims under a variety of conditions, thus helping to improve reliability and increase customer satisfaction. »

Toyota Motor Corp.

- « Toyota Motor Corp. overtook Ford Motor Co. as the world's second largest automaker by

sales volume in 2004. One reason was popular cars such as the Prius gas-electric hybrid sedan. Another reason: the information systems at its North America unit. Over the past seven years, Toyota has led the industry in how it captures and mines data from every corner of its manufacturing, logistics and sales operations to gain business intelligence. Using applications from IBM's Red Brick unit, Oracle and Hyperion Solutions Corp., Toyota re-engineered its data warehouse and saved \$30 million a year in North America. With data mining, Toyota determined more efficient transportation routes and reduced the time it takes to get its vehicles to customers from 37 days to 18. Dealers also get immediate information on vehicles they've ordered. Toyota identifies sales trends faster, and can determine the best locations for new dealerships. Research firm IDC calculated the overall project's return on investment at 506 percent. »

Herlitz AG

- « Using information gained about products, locations and specific regional characteristics, the company is helping its partners to specifically analyze the buying habits of their respective customers and find the best location for a specific item....The company derives this knowledge by analysing the reputation and range of products available, as well as regional differences. The main question here is: which factors result in one item being sold more than another?... Ultimately, the decision was made in favor of SPSS' Clementine tool...the first step Herlitz AG took to provide a range of products tailored to a specific location was to focus on preparing the data. This involved cleaning up data from the internal datawarehouse and enhancing this data with external third-party data....The datasets comprised an average of 4,000 cases with approximately 30 variables....On the basis of these results, at least one rule or profile, indicating when sales are above or below average, was generated for 68 percent, or 4,420 of the 6,441 items analysed. Using a tool developed by Herlitz, a range of products was then specified for each individual subsidiary which, taking into account all the special features of the location in question, would guarantee improved profitability. »

R.R Donnelly & Sons

- « For years, a problem that has plagued the printing industry was also affecting our rotogravure printing process. Over a given year, we were experiencing in the neighborhood of 400 to 600 cylinder bands, a phenomenon that causes streaks on the printed material. Each time this occurred, the grooves had to be eliminated by polishing, either on the spot or once the cylinder was removed. Machines would have to be shut down, thereby losing man hours and productivity was being affected. We decided the Decision Tree Induction process was a very effective means of solving this previously unsolvable problem... In the case of cylinder bands, we've saved over 800 hours annually at the Gallatin plant. We have since used KnowledgeSEEKER for gaining insights and new understanding into how to solve cylinder band problems at other plants and have achieved results equal to those at the Gallatin plant. »

Halliburton Energy Services

- « Halliburton Energy Services supplies products, services, and solutions for oil and gas exploration and production worldwide – from the initial evaluation of potential sites to drilling and well maintenance. Oil well construction begins with the drilling of a well bore. Steel casing is then inserted into the bore hole and cemented in place. To allow the oil into the bore the steel casing, the cement, and the surrounding oil-bearing formation are perforated by means of explosive charges from a perforating gun. For safety reasons, it is vital to know whether all the explosives have detonated before the detonating apparatus is brought back to the surface and removed from the well....The signals that indicate successful detonation of explosive charges are often difficult to hear because of the depth of the well bore, which can be two or three miles. In an effort to strengthen the detonation signal, Schultz [an Halliburton's research

engineer] designed a system in which accelerometers (sensors) attached to the well head capture and amplify the acoustic stress waves that travel up the pipe when the perforating guns go off. He found, however, that the sensitive accelerometers also captured signals from pumps, generators, and other equipment around the well head. Schultz needed to develop a filter that would separate the accelerometer signal from contamination caused by these ambient sounds....Working in MATLAB, Schultz developed an adaptive, predictive nonlinear neural network filter that cleanses the signals of the contaminating repetitive noises, leaving only the impulsive components – which include the signal generated by the subsurface explosion....He based his neural network code on models included in the Neural Network Toolbox^[31]....Following successful trial tests, the adaptive neural network filter is being used as the basis for other projects using adaptive neural networks, and Halliburton has initiated patent protection for the technology. »

Compaq

- « Founded in 1982, Compaq Computer Corporation is a leading global provider of enterprise technology and solutions....The two areas where SAS Enterprise Miner has helped us solve particular issues are in customer segmentation and predictive modelling. For customer segmentation, we have typically used the disjoint clustering capability and/or self-organizing maps in order to segment customers or prospects into groups of similar demographics for the purpose of CRM-targeted communications. For predictive modelling, we have used Enterprise Miner to predict customers who are similar to our best set of customers within a market segment. We use these predictions for targeted campaigns. Another application of predictive modelling is for corporate strategic business intelligence. A model was developed that estimates the IT spending of customers or prospects, and the scored results in our customer data warehouse or in the D&B database allowed us to aggregate the estimates in any custom fashion for regional or vertical market analysis....Depending on the project, the ROI ranged anywhere from two to 50 times....Another feature we will use in the near future is text mining for analysing telesales notes, e-mail text, and customer and prospect Web documents for additional business intelligence applications. »

Bayernwerke AG

- « Bayernwerke AG, located in Bavaria, Germany, is one of the largest utility in Germany. Precise Load Forecast solutions using ANN are required for optimizing their energy management system....The idea was to improve the precision of the forecast by applying neural network technologies and create a modular system. The forecast is a short-term prediction which makes possible predictions for the next 24 hours on a hourly basis (or more frequently if data is available). The forecast can be done for each day up to three weeks in the future. When compared to conventional methods, this method based on neural networks and modular system architecture performs the best. The prediction error ranges between 1 - 2% whereas multiple regressions, for instance, gives around 2 - 3% error. The difference in terms of absolute power is in the terms of 10 MW. Furthermore, the time required for calculating a forecast is less than one second and forecasts for special days (legal and religious holidays) are possible. »

John Deere Waterloo Works

- « Equipment manufacturing is an inherently capital intensive business and knowing the direction of the market three months sooner can save tens of millions of dollars in inventory costs every quarter, as well as better focusing marketing campaigns and sales incentives....By using Information Discovery pattern-based approach to data analysis, Deere can now better forecast tractor sales. The data mining approach used by Information Discovery produces much better forecasts...Good forecasts allow Deere to better manage its inventory, and make significantly better strategic marketing decisions....Sales patterns are automatically identified

and stored in a Pattern Warehouse as well as the impact of industry factors such as cash receipts by farmers, crop-future contract prices, etc. The pattern base is incrementally updated...is only a fraction of the size of the database, allowing Deere to store many years worth of patterns with ease, providing Deere with an environment for long term corporate knowledge management. »

Tauernkraftwerke AG

- « Tauernkraftwerke AG, located in Salzburg, is Austria's largest provider of hydro dam generated electrical energy. A water dam is a subject to deformation due to various influences including the height of the water level and temperature. In order to assess the security of the dam, it is important to model the displacement of the dam and to monitor the deformation by comparing measured data to the model. In case of significant differences between measured and modelled displacement, an investigation of the dam has to be carried out. The objective of this project has been to develop a model to predict the deformation on the basis of a neural network. The neural network method provides a significant improvement in the forecast compared to conventional statistic approaches (e.g. multiple regression). Depending on the forecast period, the absolute error for the neural network forecast model lies between 0.1% and 1.0%. In comparison, the conventional multiple regression based model achieves absolute errors of 2% - 3%. The neural network method supplies a precise forecast which reduces unnecessary investigations of the dam and thereby reduces operational costs. »

Southeastern US Electric Utility

- « Issues of customer satisfaction and customer retention are becoming increasingly important in today's competitive business climate. Many companies conduct regular surveys of their customers to measure how they rate the products or services delivered, customer service, reliability, etc....Z Solutions^[32] analysed a customer satisfaction survey conducted by an electric utility headquartered in the Southeast. The goal was to identify those specific aspects of service that most impacted a customer's overall satisfaction with the company, and to determine whether these key aspects had changed over time....Neural networks were successfully used for the analysis because of their ability to measure multiple and complex interactions between variables....The neural network was able to discern between different overall ratings of satisfaction by customers, and by utilizing data mining techniques the key predictors of satisfaction were extracted from the models. The key predictors analysis revealed aspects of service that customers found most important, company communication channels that were most effective, and demographic segments that were most likely to be highly satisfied. This important information was then used by company decision makers to direct marketing strategy related to customer satisfaction and retention. »

Northeastern US Electric Utility

- « A large electric utility in the Northeast needed to quantify the benefits of existing programs designed to reduce the frequency of power outages on its system. The programs may be something as simple as trimming the branches of trees near power lines, or as involved as a re-engineering of the electric distribution system. How much impact do these programs have on reducing power outages and what expenditures are cost effective? This is a difficult question to answer because outages are predominantly determined by weather conditions in the utility's service territory. A major storm will cause outages even in well-run systems....Accurately the impact of these storms based on their severity and other factors was the key challenge in the project...we [Z Solutions] collected a diverse set of data including: weather data, geographical description of service territory, customer information, time of year, outage reduction programs. All of these factors interact to determine if an outage occurs on a given day...neural network techniques were successful in processing all the factors and complex conditions involved to

arrive at a reasonable prediction of outages. The neural networks were able to strip away the complications of weather that was making program evaluation previously impossible. Therefore, the client could determine which of the programs was most effective, and budget for the optimal level of expenditures for each program. »

Miscellaneous

EDF Energy [*\[v\]*](#)

- « EDF Energy has reduced customer bad debt by 60 per cent after using data mining software to analyse its customers' financial performance. It has also used the technology to identify 1.7 million potential clients, by evaluating its criteria for new subscribers against data on UK householders. Clifford Budge, EDF Energy customer insight manager, says the Clementine data mining software from specialist vendor SPSS has become an essential tool for examining data. 'There is definitely demand to have more insight into the business,' said Budge. 'We use Clementine to look at customer retention, product information and customer acquisition.' The customer insight team was most recently asked to re-examine the rules governing the system, which scours individual customer records for common occurrences of bad debt. 'With any piece of work like this, we are looking at a problem spread over five million customer records,' said Budge. 'When we were asked to re-evaluate these rules, the model we came up with was 60 per cent better than it was before.' By predicting which geographical areas are most likely to have a high concentration of unpaid bills, the new rules are helping EDF to cut 60 per cent of the bad debt that would otherwise be wiped from its balance sheet....Similar work to improve the targeting capabilities of the sales and marketing department has increased the potential sales pool by more than 25 per cent....'The work we do is equivalent to creating 10 models a year at about £14,000 per model, where other companies can pay as much as hundreds of thousands of pounds for just one,' said Budge.» »

Anderson Analytics [*\[vi\]*](#)

- « It all started when Anderson Analytics Managing Partner and Founder, Tom Anderson, was looking for a means to compare different diamonds to each other, as well as different diamond retailers to each other, to see which diamonds were a better value for the price. "I thought that diamond retailers must have a formula for valuating diamonds and setting the correct selling and buying price based on what they call the 4 C's (cut, color, clarity and carat)," said Anderson. "But diamond merchants never reveal their methods for determining value, and to my surprise there was no such formula anywhere on the Internet." So Tom turned to Anderson Analytics to do what it does best -- take large amounts of data, such as general diamond pricing and 4 'C' information for thousands of diamonds, and analyze it. Using advanced statistical software Anderson Analytics literally 'mined' all the data available on over 44,000 diamonds! The result of the multivariate analysis was a methodology for predicting the price and market value of diamonds with a very high degree of certainty. »

Ohio Department of Natural Resources

- « In addition to resource management, ODNR is responsible for promoting leisure services and recreational opportunities for the public...."We began to realize that we needed to think about our business in a new way," said Mike Costello, program administrator of the Ohio Division of Wildlife Licensing fees support the cost of maintenance operations...."With hunting and fishing licenses on the decline, we knew we had to change from a product- to a customer-focused business," explains Costello....A shift in business philosophy was leading

ODNR toward a customer relationship management approach. Costello's team began to explore the benefits of data mining. With the recent automation of its licensing processes, ODNR has gained the ability to examine enormous amounts of previously unavailable data. From the first and second years of data collection, ODNR identified a 50 percent churn rate. Of all the people who bought fishing or hunting licenses the first year, only half came back. The agency lost 350,000 customers but gained an amazing 325,000 new ones. ODNR's CRM strategy quickly became focused on maintaining the loyalty of current customers. Predictive modelling helped assess which customers were more likely to lapse, and the agency created marketing campaigns, including postcards and ads, to strengthen customer commitment and bring them 'back to the woods.' As a result of its CRM strategies, ODNR generated more than half a million dollars in direct licensing revenues. "Convincing management to buy the technology soon became a no-brainer," said Costello. "Our return on investment was clearly evident and quickly attainable." »

State of Texas [*\[viii\]*](#)

- « The state of Texas has harnessed SPSS predictive analytics software to enhance its "Advanced Database System (ADS)" tax compliance function. Predictive models are a key component of ADS and played an important role to recover over \$400 million in unpaid taxes since its inception in 1998. The Audit Division of the Texas Comptroller of Public Accounts (CPA) uses SPSS and Elite Analytics, LLC, a data mining consulting service, to maximize taxpayer compliance and maintain revenue streams. State tax agencies are charged with reducing the "tax gap" between the tax owed and the amount collected. Audits are critical to enforcing tax laws and helping tax agencies achieve revenue objectives. With SPSS' data mining workbench, Clementine, Texas CPA and Elite Analytics developed an audit selection strategy that more accurately predicts which audit leads are more likely to yield greater tax adjustments. "State tax agencies have limited staff and resources, and predictive analytics enable the agencies to more efficiently and effectively identify delinquent taxpayers and make effective resource allocation decisions," said Daniele Micci-Barreca, PhD, a principal at Elite Analytics. "With SPSS predictive analytics, the Texas CPA was able to refine its traditional audit selection strategies to produce more accurate results." » »

Waco Police Services

- « In the heart of Texas, analysts with Waco Police Services use SAS to predict criminal activity and crime patterns. Based primarily on crime reports and histories, the Waco crime model is used to create weekly reports that map crime by area and predict the city's top 10 "hot spots." "The crime model complements our ability to predict and track crime," says Sgt. Dennis Kidwell, bureau chief for the crime analysis section, "because we're able to analyze and track a large number of observations simultaneously." Those observations include the time, date and location of all crimes, broken down into 11 major classifications, including residential burglary, vehicle burglary, criminal mischief, homicide and assault....Using SAS for analysis and software from MapInfo for GIS mapping, Waco Police Services has become more effective in preventing crime. Says Kidwell, "The crime model has been highly accurate in predicting the occurrence of criminal activity, which allows us to be more responsive in assigning the manpower to counter criminal activity." » »

Center Parcs

- « Center Parcs is Europe's market leader in the field of short vacations....The most important challenge for Center Parcs is how to reach maximum occupancy for the vacation parks. This is also called yield management. The average occupancy of the approximately nine thousand

bungalows in Europe is about 90%, which is unequalled in the leisure industry. Center Parcs believes that there is still room to increase the occupancy rate and the profitability. Center Parcs' primary marketing challenge was to decrease the number of mail-packs and to optimally target the customers most likely to respond, thus realizing a higher occupancy rate with lower costs....Center Parcs has implemented DataDistilleries' analytical Customer Relationship Management (aCRM) to optimise the yield. The advanced Customer Behavior Modeling technology in DD Series gives Center Parcs accurate and up-to-date insight into their customer base. Center Parcs has divided its customer base into four segments. DataDistilleries' role in achieving maximum occupancy is to optimise Center Parcs' direct-mail and brochure channel towards the four segments. DD Series targets customers that are likely to book a vacation for arrival within a certain arrival window, based on their past behavior, and to predict response rates for these marketing campaigns. »

Cabrillo College

- « In order to better uphold its mission, Cabrillo wanted to determine which students were most likely to drop out in order to improve student retention by offering a more relevant selection of classes scheduled at convenient times. With SPSS Inc.'s Clementine[®] as its data mining solution, Cabrillo College is gaining a deep understanding of student enrolment patterns and tendencies....Clementine allows Cabrillo to explore and evaluate a range of variables and predict each student's probability of completing a class, transferring out of a class or leaving the school altogether, "By predicting which students may need some attention or reinforcement of their education, we can provide each student on an individual basis with relevant information or discuss how we might be able to help them overcome the obstacles negating their staying," said Dr. Luan. Cabrillo used a combination of both segmentation and clustering techniques to establish typologies and to understand grouping dynamics as well as predictive modeling...."We can adjust our curriculum to add programs or subtract classes that are disadvantageous to our student's learning," said Dr. Luan. "Moreover, we can determine what classes should be offered at what times. For example, we found students with particular profiles were most likely to take night classes. Rather than print and distribute hundreds of class catalogues, with the information we obtained using Clementine, we were able to adjust class schedules and match them to students' preferences," added Dr. Luan. "This reduced our marketing budget while increasing our effectiveness." »

West Midlands Police Department

- « While many cases lacking evidence were filed away, the department is now re-examining them, and doing it more quickly than ever before. In Clementine, Adderley uses two kohonen networks to cluster similar physical descriptions and Mos [modus operandi]. He then combines clusters to see whether groups of similar physical descriptions coincide with groups of similar Mos. If he finds a good match, and perpetrators are known for one or more of the offenses, it is possible the unsolved cases were committed by the same individuals. Adderley's analytical team further investigates the clusters, using statistical methods to verify the similarities' importance. If clusters indicate the same criminal might be at work, the department is likely to re-open and investigate the other crimes. Or, if the criminal is unknown but a large cluster indicates the same offender, the leads from these cases can be combined – and the case reprioritized. Adderley is also investigating the behavior of prolific repeat offenders, with the goal of identifying crimes that seem to fit their behavioural pattern. »

State Revenue Agencies (MA, CA, TX)

- « State revenue agencies across the nation are hunting for tax evaders with new high-tech

tools: computer programs that mine an increasing number of databases for clues on the finances of people and businesses....The tax agencies' "data warehouses" can stockpile data from state and federal agencies and, in some cases, private sources. And they are using new tools to analyze the data, including "data-mining" software that can scrutinize mountains of information to find patterns or establish relationships....The Massachusetts system mixes databases from the IRS and Customs, along with state motor vehicle, incorporation and professional licensing records. The state tax agency says it uses other databases, but won't name them....The Massachusetts agency has brought in \$47 million thanks to the system since its June 2002 inception, LeBovidge [Massachusetts Revenue Commissioner] said. California officials estimate that for the four years ending in fiscal 2003, their new system brought in \$260.6 million -- while Texas says its data-mining tech has harvested more than \$362 million since the late 1990s. As an example of a successful case, Massachusetts officials said IRS records led them to a man who worked in the state but had not bothered to file state income taxes. He had to cough up \$33,000. »

Cox Communications

- « Cox Communications, a Fortune 500 company, is a multiservice communications company serving approximately 6.4 million customers nationwide [USA]. Prior to the September 2002 installation of KXEN Analytic Framework, Cox was using tools such as Oracle's Discoverer and SQL Navigator. These tools are not suited for building predictive models to support customer retention, acquisition and lifetime valuation which are fundamental company objectives....The product has reduced elapsed time for model creation, start to finish, by approximately 80 percent and reduced model building time from three weeks to one. By using this tool, the churn rate has reduced by a percentage point, and the company has realized the return on investment in the two months it has been in service....Predictive model generation is Cox's primary use, recognizing the output varies depending on the data input and designated objective. Cox uses these models to identify a customer's propensity to purchase products and services, desire to terminate a contract or whether a customer is a potential credit risk. »

Federal Bureau of Investigation (FBI)

- « The bureau's reorganization followed bruising revelations from field agents that their memos about suspicious enrolments of Middle Eastern men at flight-training schools and requests to investigate Zacarias Moussaoui, the alleged 20th hijacker, were overlooked. Critics can't hold Mueller accountable for those intelligence failures-he became FBI director on Sept. 3. But he's the point man now. To combat terrorism, the director says, the FBI will expand its use of data mining and financial-record and communication-analysis tools. He envisions the day when artificial-intelligence systems match data points to identify possible terrorist activity. In its March budget request, the bureau sought close to \$70 million to consolidate its investigative data warehouses, develop a secure network to share data with other intelligence and law-enforcement agencies, and implement new analytical and visualization software. »
- « Attorney General John Ashcroft recently announced [July 2002] that the Justice Department was loosening its guidelines to allow FBI agents to, among other things, dig into the vast commercial treasure house of data on consumers' buying habits, preferences and traits....While the FBI hasn't publicly specified how agents would use the data, experts say the bureau likely would employ a sophisticated technique called data mining to spot relationships in enormous amounts of data no human could possibly detect....Because potential terrorists purchase products, rent apartments and use credit cards, experts said the FBI hopes that analyzing the data would reveal patterns that could help prevent future

attacks. »

AC Milan Football Club

- « For years team doctors and coaches have looked for crystal balls that would show ACL injuries in the making, soothsayers who could hold forth on hamstrings that might blow, genies that could warn of a rotor cuff about to explode in the new hot prospect's shoulder. AC Milan may have found such an oracle: a computer smart enough to recognize the signs of an athlete coming apart. The renowned Italian soccer club – which has four players competing in this year's [2002] World Cup – has teamed up with Computer Associates International [using CleverPath Predictive Analysis Server] to test the feasibility of using neural networks, a form of artificial intelligence, to predict injuries and optimise conditioning for each athlete, perhaps even to help select which players to sign....An 18-month test of the system gave promising results....The pilot program showed that injury prediction was a possibility....Ultimately, the neural network correctly predicted injuries 84 percent of the time....The new system could be of great help to coaches trying to predict which players are at risk, said Dr. Arthur Bartolozzi, chief of sports medicine at the Pennsylvania Hospital in Philadelphia. "A lot of factors can contribute to injury: conditioning, fatigue, equipment, weather conditions, surface conditions," Bartolozzi said. If a system could be developed to take all of these things into account, it might allow team doctors to pull players before they get injured, he added....The program might also have saved AC Milan a lot of money, had it been around when the team was bidding for Fernando Redondo. "We spent an enormous amount of money buying a player who got hurt after three minutes on the treadmill," Meersseman [head of AC Milan's medical team] said. "Maybe we would have thought twice about buying him. Maybe the price would have been different." And if the team had gone through with the deal anyway, the neural network might have told them how to avoid the big injury. »

Polyphonic SMI

- « The magic ingredient set to revolutionise the pop industry is, simply, a piece of software that can "predict" the chance of a track being a hit or a miss. This computerised equivalent of the television programmer Juke Box Jury is known as Hit Song Science (HSS). It has been developed by a Spanish company, Polyphonic HMI...It isolated and separated 20 aspects of song construction including melody, harmony, chord progression, beat, tempo and pitch and identifies and maps recurrent patterns in a song, before matching it against a database containing 30 years' worth of Billboard hit singles - 3.5m tunes in all. The program then accords the song a score, which registers, in effect, the likelihood of it being a chart success....HSS confidently predicted Norah Jones's meteoric success (tipping no less than 10 songs on her debut album *Come Away with Me*) well in advance of her chart-topping appearances and in the face of an industry unconvinced she would have any commercial impact....Of course, the appeal to record labels is obvious, as it offers a rational underpinning for commercial decisions. With the recordings themselves being the least expensive element of launching an act, the marketing resource being the greatest, and most companies being run by bean counters, we can be certain that this kind of analytical software won't go awayIt's all in the clusters, you see. Hit songs, typically, fall into one of a number of groupings - there are around 50 in the US and 60 in the UK where, traditionally, tastes have been more diverse. Belonging to the same cluster does not mean songs sound the same, though, more that they are mathematically similar. And the analysis has thrown up some very unlikely musical bedfellows: Some U2 songs are in the same cluster as Beethoven, while spandex ultra rocker Van Halen sits right alongside MOR piano babe Vanessa Carlton. It is for this reason that Polyphonic are confident their software won't homogenise our already stratified and similar sounding charts. They are already working

with one radio station to expand their playlist without losing audience share by selecting songs with the correct mathematical rhythms. In a world where drearily repetitive playlists have become the norm this could be the answer to an oft-uttered prayer. This strategic approach may seal the software's place in history. McCready [CEO of Polyphonic] explains how they are helping a very well known "smooth male jazz crooner" who is finding it difficult to break into the US market. The label's marketing department are promoting him to the Norah Jones audience. But Polyphonic's analysis has shown that the crooner's song patterns are more similar to Linkin Park, Aerosmith and JayZ. This kind of interpretation offers an unprecedented rationale for appealing to a seemingly unlikely demographic....Ric Wake, producer of international acts such as Jennifer Lopez and Anastacia, has drawn the technology into the heart of the creative process. When you're only a few "mathematical rhythms" away from a great hit this could save hours, days, even weeks of studio grind. At the end of each day relevant tracks are downloaded and feedback is presented the next morning. Supporters of the software argue that it does not detract from the artistic process; it is still the humans who must find the solutions to a low-scoring song. »

Hungarian Museum of War History

- « In 2001, the Institute of Hungarian History of War purchased the records of those who died in the camps from the Hungarian government and began working on them. The researchers found that because the Soviet record-keepers had used Cyrillic letters rather than Latin, and because much of the most vital information (such as Hungarian names and cities) was missing, traditional technological solutions could not be used to match the Soviet records to the Hungarian military records....By chance, Zoltan Benedek heard about the data quality problems and had an idea. Benedek, a consultant in the data mining division of KFKI ISYS Information Systems Ltd., contacted János Bús, director of central archives at the Hungarian Museum of War History, with an offer to help....KFKI ISYS staff members recognized that the problem facing the institute was similar to one they themselves had faced while consolidating databases and building SAS data warehouses for clients. In these situations, various departments and systems referred to specific records by several different names. In addition, numerous human errors and missing values made it a challenge to sort records for analysis. To formulate a solution, they relied on the power of SAS software and a data mining technique called "textual link analysis."....The project did not take long for the data mining team at KFKI ISYS to complete, and they were able to publish the search results on the Internet. The cleaned data sets brought to light information on approximately 27,000 people who had disappeared during the war. The first person identified had lived five years longer than stated in official Hungarian government records. He had been imprisoned for five years and had worked in a Soviet mine several thousand kilometres away from his home after his family believed him to be dead....THE KFKI ISYS team is receiving e-mails daily from grateful Hungarians all over the world who are finally able to discover information about their relatives' remains and exact dates of death. »

FineArtExplorer.com

- « Discovering art you like just got a lot easier. And for artists, being discovered is now a lot easier as well. Z Solutions, an Atlanta based developer of adaptive learning technology, has announced the launch of its FineArtExplorer website (<http://www.FineArtExplorer.com>) powered by the MyMuzeä Discovery Engine....The FineArtExplorer site is of value to both artists and art buyers. Artists don't have to describe or characterize their art in words to be found and enjoyed online. Art buyers are directed to works that match their taste without extensive searching. FineArtExplorer.com is like hav[ing] a shopping adviser who learns your taste then displays only items of interest. Users are walked through a gallery of fine art online. As each image is rated by the viewer, a profile of the user's taste is created. After

surveying two dozen pieces, the discovery engine goes to work and creates a personal gallery of art that it recommends. The user can then click through to the host site of the artist and learn more, see more or purchase art. The personal gallery can be saved and is updated as new art matching the profile is discovered. »

DynMeridian

- « DynMeridian, a DynCorp company, is a professional services firm that provides comprehensive analytical and technical support in arms control, national security affairs and related high technology to U.S. government and industry clients....The project involved exploring recent technological advancements in data mining and XML to determine the feasibility of leveraging these techniques in support of early identification of and/or prediction of personnel retention trends, as well as the efficient transfer of the information to the decision makers. PolyAnalyst was used to data mine the personnel database to generate insightful information and entity-level propensity-to-lose results....In this project, scoring test data with the developed classification model produced results where the top 10 percent of the cases of the predicted most loyal personnel contained over 60 percent of all people who indeed served for a long time. This represents a greater than six-fold lift and allows the client to better target expensive personnel loyalty programs to the most qualified candidates. »

Northwood Inc.

- « Northwood Inc. is a major integrated forest products company located in Prince George, British Columbia. Northwood operates five lumber mills, a pulp mill, a plywood mill and a treated wood plant...The company also operates a forest center and nursery, which produces approximately eight million seedlings a year... For a young seedling vegetation can be deadly. In order to protect its investment, Northwood had to apply for herbicide permits on a case-by-case basis to control vegetation and protect newly planted trees. In 1997, the government of British Columbia changed the law so timber companies could submit a pre-planned, integrated pest management plan for approval, eliminating the need for individual permits. Before developing the plan, Northwood had to determine under which conditions a tree was threatened by vegetation and when it was necessary to apply herbicide. In order to accurately define standards and thresholds for vegetation management, Northwood had to identify patterns among the numerous variables in a database of 15,000 trees, containing data collected over nine years....Northwood used Cognos Scenario^[33] to identify patterns of tree mortality and growth and rank critical factors in a tree's ability to reach maturity. Based on the information mined from the data through Scenario, Northwood was able to outline at what thresholds tree were under-performing. With this knowledge, Northwood drew up an integrated pest management plan that was approved by the government of British Columbia. Northwood is now better able to manage vegetation, while using less herbicide and seeing more trees reach full growth. »

East Ayrshire Council

- « Due to the introduction of the 4 C's of the Best Value Regime, challenge, consult, compare and compete, all Councils must demonstrate how they consult citizens, service users and others for their views....Since 1999, East Ayrshire Council have developed and distributed a number of different questionnaires in order to support their Best Value initiatives....Since every Department within the Council must consult with citizens, the reliance on a manual system started to cause problems. One of the major problems was that by the time the information was analysed, it was almost out of date. Analysts at East Ayrshire Council held the view that they needed a more modern approach to replace the existing pen, paper and graphs in Excel methodology....By adopting an analytical solution from SPSS, East Ayrshire

Council were more able to deliver better service to their citizens....By using SPSS Base and Data Entry software to code, input and analyse questionnaires, East Ayrshire Council have saved a huge amount of staff time whilst also improving the accuracy and quality of the results. »

MetLife Insurance Company

- « While there appears to be no shortage of applications that are candidates for conventional auditing tools such as Caats, there have not been many such examples for the ‘exotic tools’ such as data mining and data visualization software. This situation should change dramatically as auditing units start to achieve some success in using these tools. Human Resource Applications represent one of the virgin areas for data mining....WizRule will indicate anomalies based upon deviations from salary averages for specific levels. While these may or may not point to problem area, auditors should satisfy any concerns that may have related to the deviations....If you include the distance from the home to the office or time to travel this distance as a variable, WizRule will determine any variance from the norm. While on the surface this may appear to be an unimportant finding, it can point out some very serious areas that should be of concern to the auditor. Employees who have to travel great distances to get to work are good candidates for excessive lateness, absences and eventually leaving for a position closer to their residence....Employees, who do not participate in 401K matching programs, medical benefit programs, etc., may be trying to hide something. While this is not always the case, it is area worth investigating. (E.g. the employee is receiving two payroll checks!)....As an example, WizRule will indicate that 67 of 68 employees in the Detroit office belong to the Central Claims group. One employee belongs to the Southern Claims group. (Again a possible ghost employee)....These are just a few examples that can be employed using Data Mining Tools in most Corporate Human Resource Systems. »

US Department of Agriculture

- « Commercial lenders use the technology to predict loan-default or poor-repayment behaviors at the time they decide to make a loan. The USDA’s principal interest, on the other hand, lies in predicting problems for loans already in place. Isolating problem loans lets the USDA devote more attention and assistance to such borrowers, thereby reducing the likelihood that their loans will become problems....The USDA retained my company, Exclusive Ore^[34], to provide it with data mining training. As part of that training, my colleagues and I performed a preliminary study with...a small sample of current mortgages...roughly 2 percent of the USDA’s current database....At the USDA, our goal was to build a model that would predict the loan classification based on information about the loan, borrower, and property....For the USDA, the initial models revealed that the important factors to loan outcome included loan type, such as regular or construction; type of security, such as first mortgage or junior mortgage; marital status; and monthly payment size....The USDA’s preliminary data mining study sought to demonstrate the technology’s potential as a predictor and learning tool...the department plans to expand the limited number of attribute available...to include payment histories....Eventually, the USDA will use these models to identify loans for added attention and support, with the goal of reducing late payments and defaults. »
- « Agriculture Department officials were alerted to more than \$250 million in fraudulent crop insurance claims in the past three years after they began using data mining. Based on that initial success, officials in the USDA's Risk Management Agency are exploring additional uses of data mining for improving Federal Crop Insurance Corp. policies and procedures....Since 2000, when lawmakers allocated \$20 million for a five-year study to reduce waste, fraud and abuse in the insurance program, USDA officials have documented a

better than 20 to 1 return on the \$13 million they have invested in data mining....In the past few years, agency officials paid about \$3 billion a year for legitimate crop loss claims, Westmoreland [director of strategic data acquisition and analysis at the agency] said. But data mining has shown that about 1,800 out of 1.5 million people enrolled in the crop insurance program attempt to file fraudulent claims. Before officials discovered data mining, the USDA typically lost about \$7 million a year from fraudulent claims. Officials tried to recover the money, but litigation is difficult and time-consuming, Westmoreland said. "We would rather people make sure they file a valid claim and get paid for a valid claim than try to rectify claims that may have been invalid," he said. "It just works out better for everyone involved." »

Compaq Computer Corp.

- « Sales and marketing executives at Compaq Computer Corp. count on text mining tools to analyze company descriptions in their prospect database. The results help executives target customers for new sales and marketing campaigns. »

University of Louisville Medical Center

- « A new text mining project at the University of Louisville Medical Center will let doctors make better use of medical databases such as Medline, PsychInfo and Toxline for evidence-based medicine. Search results of these medical databases can often yield 2,000 matches, but advanced modelling with Enterprise Miner can reduce the results to 100 highly relevant documents and sort those 100 documents into smaller subgroups or categories. »

[1] An earlier version of this document appeared as Data/Text Mining Case Studies Repository, An ELCA Informatique SA White Paper, 2004

[2] Clementine is SPSS Inc.'s Data Mining tool. See <http://www.spss.com> for details.

[3] See <http://www.datadistilleries.com> for details of DataDistilleries' range of analytical CRM products.

[4] See <http://www.ca.com> for details of Computer Associates International's analytical products and services.

[5] DataScope is Cygron Pte. Ltd's Data Mining tool. See <http://www.cygron.com> for details.

[6] WizRule is WizSoft, Inc.'s data auditing and cleansing application, specialised in anomaly/deviation detection. See <http://www.wizsoft.com> for details.

[7] Intelligent Miner is IBM's Data Mining tool. See <http://www.ibm.com> for details.

[8] ModelMAX is ASA's Data Mining tool. See <http://www.asacorp.com/products> for details.

[9] DigiMine Services is a fully hosted data warehousing and data mining solution which provides advanced analytics to e-businesses. See <http://www.digimine.com> for details.

[10] Enterprise Miner is SAS Institute's Data Mining tool. See <http://www.sas.com> for details.

[11] Quadstone System is Quadstone, Inc.'s Data Mining tool. See <http://www.quadstone.com> for details.

[12] KXEN Analytic Framework is KXEN's suite of Data Mining tools. See <http://www.kxen.com> for details.

[13] WizWhy is WizSoft, Inc.'s Data Mining tool. See <http://www.wizsoft.com> for details.

[14] See http://www.businessobjects.com/applications/aa_products.htm for details of Business Objects' analytical products and services.

[15] KnowledgeSEEKER is Angoss Software Corporation's original Data Mining package. See <http://www.angoss.com> for details.

[16] See <http://www.data-miners.com> for details.

[17] KnowledgeSTUDIO is Angoss Software Corporation's advanced Data Mining tool. See <http://www.angoss.com> for details.

- [18] See <http://www.fairisaac.com> for details of Fair, Isaac's analytics offerings.
- [19] ScorXPRESS is ASA's customised Data Mining solution for fraud detection. See <http://www.asacorp.com/products> for details.
- [20] See <http://www.teradata.com> for details of Teradata's DWH and DM products and services
- [21] PolyAnalyst is Megaputer Intelligence, Inc.'s Data Mining tool. See <http://www.megaputer.com> for details.
- [22] BusinessMiner remains part of Cognos' BusinessObjects package. It is however being extended by new products and services.
- [23] Recon is a trademark and service mark of Lockheed Martin Missiles and Space Company, Inc. Further details are in relevant papers by G.H. John (see <http://robotics.stanford.edu/~gjohn>)
- [24] CART is Salford Systems' Data Mining tool. See <http://www.salford-systems.com> for details.
- [25] See <http://www.polyvista.com> for details of PolyVista's suite of analytics software.
- [26] See <http://www.mathworks.com> for details of MATLAB's extensive range of statistics and data analysis techniques, in particular its Neural Network Toolbox.
- [27] S-PLUS is Insightful Corporation's (formerly MathSoft, Data Analysis Products Division) data analysis tool. See <http://www.insightful.com> for details.
- [28] This work was recognized by Computerworld Honors Foundation for Outstanding Achievement in Medicine and is now part of the Computerworld Honors Foundation's collection of case studies (20044978) and the Smithsonian Institution's permanent collection.
- [29] See <http://www.neuroshell.com> for further details on the NeuroShell product range.
- [30] Text Miner is SAS's text mining tool, part of Enterprise Miner. See <http://www.sas.com> for details.
- [31] The Neural Network Toolbox is one of MATLAB's specialised toolbox for advanced data analysis. See <http://www.mathworks.com/products/neuralnet> for details.
- [32] See <http://www.zsolutions.com> for details of Z Solutions' Data Mining software and consultancy services, specialised in neural networks.
- [33] See <http://www.cognos.com/products/datamining.html> for details of Cognos' Data Mining products and services.
- [34] Exclusive Ore, Inc. is a data mining and database management consultancy. See <http://www.xore.com> for details.

-
- [i] Leading Danish Insurer Improves Claim Handling & Customer Support with SPSS Predictive Analytics, KDnuggets News, No. 4, Item 40, 2006 (www.kdnuggets.com/news/2006/n04/40i.html)
- [ii] Pediatrix Data Mining Identifies Favorable Antibiotic Combination for At-Risk Infants, Genetic Engineering News, Breaking News, 27 January 2006 (www.genengnews.com/news/bnitem.aspx?name=1153932XSL_NEWSML_TO_NEWSML.xml)
- [iii] Data Mining: Solving Care, Cost Capers, Greg Gillespie, *Health Data Management*, January 2005 (www.healthdatamanagement.com/html/current/PastIssueStory.cfm?ArticleId=10300&issuedate=2004-11-01)
- [iv] Baylor Health Care System Implements SAS Software to Evaluate Technology Efforts, DM Direct Newsletter: Industry Implementations, DMReview Web Editorial Staff, 8 April 2005 (www.dmreview.com/editorial/newsletter_article.cfm?nl=dmdirect&articleId=1025098&issue=20167)
- [v] EDF Energy keeps eye on bad debt, Miya Knights, Computing, 10 November 2005 (www.vnunet.com/computing/news/2145830/edf-energy-keeps-eye-bad-debt)
- [vi] Anderson Analytics: Data Mining for Diamonds, CRM Today, 13 February 2006 (www.crm2day.com/news/crm/117346.php)
- [vii] SPSS Predictive Analytics Helps State of Texas Recover \$400 Million in Unpaid Taxes, DM Direct Newsletter: Industry Implementations, DMReview.com Web Editorial Staff, 8 July 2005 (www.dmreview.com/article_sub.cfm?articleId=1031670)