Cornell Hotel and Restaurant Administration Quarterly

Data Mining for Hotel Firms: Use and Limitations

Vincent P. Magnini, Earl D. Honeycutt, Jr. and Sharon K. Hodge

The online version of this article can be found at:

Published by:

$SAGE

On behalf of:

Additional services and information for *Cornell Hotel and Restaurant Administration Quarterly* can be found at:

Email Alerts: http://cqx.sagepub.com/cgi/alerts

Subscriptions: http://cqx.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Data Mining for Hotel Firms:

Use and Limitations

Data mining can provide a window into customers' behavior—if it's handled correctly.

BY VINCENT P. MAGNINI, EARL D. HONEYCUTT, JR., AND SHARON K. HODGE

In the hotel industry, knowing your guests—where they are from, how much they spend, and when and on what they spend it—can help you formulate marketing strategies and maximize profits. Fueled by the proliferation of centralized reservation and property-management systems, hotel corporations accumulate large amounts of consumer data. This information can be organized and integrated in databases that can then be tapped to guide marketing decisions. However, identifying important variables and relationships located in these consumer-information systems can be a daunting task. The relatively new process known as data mining can be instrumental in overcoming such obstacles.[1] From stores of information, data-mining technology extracts meaningful patterns and builds predictive customer-behavior models that aid in decision making.[2]

Data mining is a largely automated process that uses statistical analyses to sift through massive data sets to detect useful, non-obvious, and previously unknown patterns or data trends.[3] The emphasis is on the computer-based exploration of previously uncharted relationships (i.e., using "machine learning" methods that typically require only limited human involvement).[4] Without data mining, valuable marketing in-

---

[1] For a discussion of the use of compiled data, see: Robert K. Griffin, "Data Warehousing: The Latest Strategic Weapon for the Lodging Industry?," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 39, No. 4 (August 1998), pp. 28–35. For a discussion of the use of guest-history data, see: Paula A. Francese and Leo M. Renaghan, "Data-base Marketing: Building Customer Profiles," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 31, No. 1 (May 1990), pp. 60–63.

[2] See: A. Kamrani, W. Rong, and R. Gonzalez, "A Genetic Algorithm Methodology for Data Mining and Intelligent Knowledge Acquisition," *Computers & Industrial Engineering*, Vol. 40, No. 4 (2001), pp. 361–377.

[3] W. Frawley, C. Piatetsky-Shapiro, and C. Matheus, "Knowledge Discovery in Databases: An Overview," *AI Magazine,* Fall 1992, pp. 213–228.

[4] P.R. Peacock, "Data Mining in Marketing: Part 1," *Marketing Management,* Winter 1998, pp. 9–18.

© 2003, CORNELL UNIVERSITY

sights about customers' characteristics and purchase patterns may remain largely untapped.[5] By uncovering such previously unknown relationships, managers have the potential to develop a winning marketing strategy that increases their hotel's bottom line.

Hotel managers understand the importance of adapting to the changing business environment not only to remain competitive, but merely to survive. As a result, technology has become a large and growing expense for many hotel corporations. Under such a technology framework, data mining is a valuable competitive tool being adopted by hotel corporations in an effort to create customer value. However, given the importance and complexity of data mining, senior hotel managers report a low level of understanding about data mining's capabilities, how it works, and what value this technology contributes.[6] The purpose of this paper is to educate hotel managers about the benefits and application of data mining on the properties they oversee.

Data Mining vs. Statistical Modeling

Data mining differs from traditional statistical modeling in a variety of ways. Data mining focuses on machine-driven model building, while statistical modeling stresses theory-driven hypothesis testing. Data-mining techniques build models, whereas classical statistical tools are supervised by a trained researcher who possesses a preconceived notion of what to examine. With statistical *a priori* analysis, relevant associations may be overlooked. By building dependency hypotheses instead of merely verifying them, though, data-mining techniques reveal important links. For example, Marriott Vacation Club International reduced the volume of direct mail it needed to reach target sales levels by correlating response rates to specific vacation offerings and specific customer characteristics.[7]

Data mining also offers enormous gains in terms of performance, speed of use, and user friendliness.[8] While data miners must understand statistical principles, highly specialized statistical knowledge is not necessary to study, understand, and improve decision-making processes. Data mining helps managers to spot trends more quickly.

Because researchers may ignore the assumptions and limitations of a theoretical model, traditional statistical analyses in customer-satisfaction research are often biased. Satisfaction research includes measures of the importance that customers place on product and services attributes. Typically, these measures are highly correlated, which can dramatically bias the statistical values that determine attributes' importance rankings. Also, statistical analyses usually assume that relationships between independent and dependent variables are linear—which is often not the case. Therefore, violation of these assumptions can result in biased and misleading statistical outcomes. Data-mining techniques (e.g., neural networks) overcome these limitations and outperform traditional statistical analyses in cases where such assumptions do not apply.[9]

Another considerable advantage over traditional statistical models is data mining's ability to easily handle large and complex datasets.[10] Data-mining techniques are not hampered by large numbers of predictive variables, and that feature makes data mining useful for selecting variables, that is, identifying those within a set that are most relevant. The ability to handle large numbers of variables also makes data mining more realistic than statistical models in representing the complexity of a typical business environment.

While many analytical techniques can be classified as data-mining tools, opinion has not coalesced regarding exactly which techniques should be considered part of the data-mining tool kit. The tools listed in the accompanying sidebar (on the next page) almost certainly belong, however.

---

[5] M. Shaw, C. Subramaniam, G. Tan, and M. Welge, "Knowledge Management and Data Mining for Marketing," *Decision Support Systems,* May 2001, pp. 127–137.

[6] C.S. Dev and M.D. Olsen, "Marketing Challenges for the Next Decade," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 41, No. 1 (February 2000), pp. 41–47.

[7] Peacock, *op. cit.*

---

[8] Le Bret, *op. cit.*

[9] *Ibid.*

[10] Peacock, *op. cit.*

A Data-mining Toolkit

- *Association rules*: Information from customer-purchase histories is used to formulate probabilistic rules for subsequent purchases.

- *Case-based reasoning*: Sets of attributes from new problems are compared with attribute sets from previously encountered problems (called cases) to find one or more boilerplate examples that provided good outcomes or solutions.

- *Decision trees*: Automatically constructed from data, these yield a sequence of step-wise rules; good for identifying important predictor variables, non-linear relationships, and interactions among variables.

- *Descriptive statistics*: Averages, variation, counts, percentages, cross-tabs, simple correlation; used at the beginning of the data-mining process to depict structure and identify potential problems in data.

- *Genetic algorithms*: Use procedures modeled on evolutionary biology (e.g., selection, mutation, survival of the fittest) to solve prediction and classification problems or develop sets of decision rules.

- *Neural networks*: Applications that mimic the processes of the human brain; capable of learning from examples (large training sets of data) to discover patterns in data; can combine information from many predictors and work well even with correlated variables, non-linear relationships, and missing data.

- *Query tools*: Provide summary measures such as counts, totals, and averages.

- *Regression-type models*: Ordinary least-squares regression, logistic regression, discriminant analysis; used mostly for confirmation of models built by "machine-learning" techniques.

- *Visualization tools*: Histograms, box plots, scatter diagrams; useful for condensing large amounts of data into a concise, comprehensible picture.—*V.P.M., E.D.H., and S.K.H.*

Looking at that toolkit, decision trees, association rules, case-based learning tools, neural networks, and genetic algorithms are categorized as machine-learning methods, while the others can be thought of as machine-assisted aids to support human learning.[11]

With data-mining techniques, levels of *a priori* specification can vary. In some cases, certain independent variables and dependent variables may be specified for examination, while predictor variables in other cases may be uncovered only by the data-mining tool. The point remains, though, that in comparison with traditional statistical methods, data-mining techniques invariably are more data driven than they are user driven.

We have observed that some hotel corporations are attempting to harness the power of information by investing in data-mining technology that exploits consumer information. Hilton Corporation uses E.piphany E.4 software at its Beverly Hills headquarters, for instance,[12] and Starwood Corporation recently invested in Unica Corp's Affinium software.[13] Such data-mining technology allows hotel corporations to predict consumer-behavior trends, which are potentially useful for marketing applications. For example, Starwood's marketing staff can run reports and analysis on customer and occupancy data stored in a data warehouse that combines customer and transaction information from all company properties. Such information indicates where customers who visit a specific hotel live. If the data reveal that the Sheraton Fisherman's Wharf in San Francisco experiences a surge in visitors from Fort Lauderdale in April, for instance, hotel marketers can increase promotional efforts in Fort Lauderdale during the late winter months.[14] Exhibit 1 lists examples of how information gleaned from data mining can be used in a hotel corporation's marketing activities.

[11] *Ibid.*

[12] L. Stevens, "CRM Analytics—CRM by the Slice—Running Analytics Is Expensive, So Companies Are Focusing on Areas with Customers," *Internetweek*, April 9, 2001, pp. 35–38.

[13] G. Tischelle and J. Maselli, "Hotels Turn to IT to Stem Losses," *Informationweek*, December 17, 2001, pp. 31–32.

[14] *Ibid.*

Harrah's Data-mining Success Story

In 1997 Harrah's hotels and casinos introduced a trademarked loyalty-card program, "Total Rewards," which tracks customers' purchasing activities and provides rewards that encourage spending at Harrah's properties. Rather than build glitzy properties with eye-popping attractions, Harrah's pursued a customer-service-oriented strategy centered around data-mining techniques. Harrah's used an information system called WINet to link all its properties, allowing the firm to collect and share customer information company wide. The process effectively changed the corporate culture from an every-property-for-itself mentality to a collaborative, customer-focused enterprise.[15]

The WINet system connects and consolidates customer information from all of the company's transaction, slot-machine, hotel-management, and reservation systems. Key pieces of information—gender, age, place of residence, and types of casino games played—help predict which customers are most likely to become frequent users. Based on this information, Harrah's designs marketing strategies to retain those customers. Customers' purchasing and gaming patterns are tracked, too, so that the company can target its customers with the most-appropriate incentives. For example, customers who reside outside the local area receive complimentary hotel rooms or transportation, while drive-in customers receive food, entertainment, or cash incentives.[16]

Data-mining techniques help to reveal data patterns and relationships that can be used to develop strong models for predicting the potential value of each customer. Given that retaining a customer is less costly than attracting a new one, building strong relationships with valued existing customers can boost profits. Having information regarding such things as the customer's birthday, anniversary, and favorite foods and drinks allows a hotel to provide excellent, tailored customer service that cements brand loyalty. Harrah's discovered that the 30 percent of its customers who spent between $100 and $500

[15] M. Levinson, "Harrah's Knows What You Did Last Night," *Darwin Magazine,* May 2001.

[16] J.A. Nickell, "Welcome to Harrah's," *Business 2.0,* April 2002.

EXHIBIT 1

Examples of the uses of data-mining information in hotel marketing

- Create direct-mail campaigns
- Plan seasonal promotions
- Plan the timing and placement of ad campaigns
- Create personalized advertisements
- Define which market segments are growing most rapidly
- Determine the number of rooms to reserve for wholesale customers and business travelers.

---

per visit accounted for 80 percent of company revenues and generated nearly 100 percent of profits. In the first two years of its rewards program, Harrah's saw a $100-million increase in revenue from customers who visited more than one property.[17] Currently, Harrah's ranks first in the industry in profit growth.[18]

Because the WINet system can consistently identify which customers will be most valuable over the long term, data mining is also useful for determining when to avoid offering incentives to customers who are not lucrative. Harrah's estimates that it has saved some $20 million by withdrawing incentives from customers who are not likely to return.[19]

Despite Harrah's success, some remain skeptical of data mining's customer benefits and long-term financial payoffs. As an example, Susan Dobscha, co-author of "Preventing the Premature Death of Relationship Marketing," advises hotels that giant central databases "are not where customers want a relationship forged. A customer would probably prefer a lower price over, say, having their beverage choice anticipated."[20]

Another important caveat regarding data mining is that any relationship discovered must be valid to benefit a company's performance. When British Columbia Telecom tried to reward 100 of its best customers by inviting them to a

[17] *Ibid.*

[18] Levinson, *op. cit.*

[19] *Ibid.*

[20] "Mining Hotel Data," *Data Warehouse Report,* October 20, 1998.

Vancouver Grizzlies basketball game, for instance, it selected customers from the database comprising frequent 900-number users. After sending invitations to the printer, the marketing staff realized that those 900-number users included a large number of sex-line enthusiasts. The company avoided a serious gaffe by refining the criteria to create a list of truly loyal guests.[21]

Data-mining Applications for the Hotel Industry

The tasks performed by data mining can be grouped into the following five categories.

(1) Classification arranges customers into pre-defined segments that allow the size and structure of market groups to be monitored. Also, predictive models can be built to classify activities. An illustration of such a model is one that predicts which segment's usage rate will experience the largest decrease when a particular promotion expires. Classification uses the information contained in sets of predictor variables, such as demographic and lifestyle data, to assign customers to segments.

(2) Clustering groups customers based on domain knowledge and the database, but does not rely on predetermined group definitions. This function is beneficial because it aids hoteliers in understanding who are their customers. For example, clustering may reveal a subgroup within a predetermined segment with homogenous purchasing behavior (e.g., a subgroup of holiday shoppers within the transient segment) that can be targeted effectively through a specific ad campaign. (The idea is that the members of the subgroup will increase their number of stays or become more loyal.) On the other hand, clustering may indicate that previously determined segments are not parsimonious and should be consolidated to increase advertising efficiency. Information such as demographic characteristics,

lifestyle descriptors, and actual product purchases are typically used in clustering.

(3) Deviation detection uncovers data anomalies, such as a sudden increase in purchases by a customer. Information of this type can prove useful if a hotel corporation wants to thank a guest for her or his recent increase in spending or offer a promotion in appreciation. Marketing managers may also attempt to draw correlations between surges in deviations with uncontrollable business-environment factors that are not represented in the database (e.g., a sharp increase in gasoline prices).

(4) Association entails the detection of connections between records, driven by association and sequence discovery. For example, a possible detected association may be that a particular segment's average length of stay increases after a specific advertising campaign. Another association task could be employed in an effort to determine why a specific promotion was successful in one market, but ineffective elsewhere. Specific information regarding customer-purchase histories is necessary to formulate probabilistic rules pertaining to subsequent purchases.

(5) Forecasting predicts the future value of continuous variables based on patterns and trends within the data. For instance, the forecasting function can be used to predict the future size of market segments. With forecasting one can also use data trends to project which hotel amenities are of growing importance to consumers and will be key drivers of the consumer's future perception of value.

In the hotel industry, the most common sources of data are CRSs and PMSs. Some hotel corporations also use information that resides in guest-loyalty-program databases. Hilton, for instance, analyzes data contained within its trade-marked Hilton HHonors database.[22] Another

---

[21] S. Press, "Fool's Gold?," *Sales and Marketing Management,* June 1998, pp. 58–61.

[22] L. Stevens, "IT Sharpens Data-mining Focus—Instead of Building Data-mining Applications with No Clear Goal, Companies Are Setting Priorities Up Front to Maximize ROI," *Internetweek,* August 6, 2001, pp. 29–30.

potentially important source of data is the information provided by guest-satisfaction surveys.

Guidelines for Effective Data Mining

When properly employed, data mining is a powerful and valuable marketing tool. However, simply investing in data-mining technology may not guarantee success. As presented below, seven guidelines influence the effective management of data-mining technology.

Guideline #1: *Match your IT priorities with an appropriate provider.* There is high demand for and low supply of data-mining expertise as more companies realize the potential value of the information residing within their databases. To capitalize on this demand, a number of second-tier research firms now provide data-mining services.[23] However, providers offer a wide range of skill levels. The most-skilled providers can turn data into useful information. Companies that initially set clear priorities have a greater chance of reaping maximum benefits from data-mining projects than do firms that are unsure of their goals.[24] Clear priorities include goals about what the firm would like to achieve through data mining and when it will be achieved. Without goals and objectives the hotel corporation is uncertain about what it is shopping for when seeking a data miner. It is also important to communicate these goals to prospective providers. When selecting a provider, ask the following six questions:

- *Does the provider have experience setting up predictive models with marketing applications?* Data mining has applications other than marketing. Data mining's ability to detect patterns in data is used extensively in criminal justice and anti-terrorism efforts to anticipate illegal activity, for instance. Wall Street also employs data mining to predict moves in the financial markets. Large global corporations use data mining to gain efficiencies in purchasing and production throughout their networks. Therefore, it is not enough to have a data-mining consultant, but one must find a provider that has experience

in marketing. Building models to predict consumer behavior is a form of data mining that requires specific expertise. For example, a data miner with marketing-applications experience would know to replace a zip code with resident characteristics, such as median income.[25]

> **Data mining can be a powerful and valuable marketing tool. However, simply investing in data-mining technology may not guarantee success.**

- *Does the provider have experience in creating models within the hospitality industry?* Marketing applications of data mining are employed across diverse industries. Building predictive models for a grocery store, a furniture chain, an airline, or a hotel is different in each case. It is beneficial to find a provider that has experience in setting up models in the hotel industry. Such a provider would more clearly understand hotel-guest-segmentation processes, for example.
- *Is the provider reputable?* Because many second-tier companies provide mining services, it is important to check the credentials and reputation of the vendor.
- *Does the provider offer the latest technology that is appropriate?* Because of the wide range of products available, it pays to do your homework. It is crucial to invest in the latest appropriate technology because it is extremely expensive and time consuming to switch products after one is installed—in no small part because switching products requires retraining the IT and marketing staff.
- *Does the provider offer a product that has visual-exploration capabilities?* Cutting-edge data-mining software has visual-exploration capabilities, which means

---

[23] M. Brandel, "Spinning Data into Gold," *Computerworld*, May 26, 2001, pp. 67–70.

[24] Stevens, *op. cit.*

[25] Brandel, *op. cit.*

Examples of hotel-guest segments

Geographic

    Nations

    States

    Counties

    Cities

Demographic

    Age or life-cycle stage

    Gender

    Income

Psychographic

    Social class

    Life-style

    Personality

    Behavior

    Occasion of purchase decision

    Occasion of use

    Benefits sought

    User status (e.g., potential, former, first time)

    Usage rate

    Loyalty status

    Buyer-readiness stage

Source: P. Kotler, J. Bowen, and J. Makens, *Marketing for Hospitality and Tourism*, second edition (Upper Saddle River, NJ: Prentice-Hall, 1999).

that data patterns can be viewed as three-dimensional objects that can be rotated or zoomed for detailed analysis. In addition, pixel-oriented technology assigns colors to data values so that patterns and trends can be examined. Visual exploration is an immense aid to managers and marketers because it often serves as a preliminary tool in selecting the appropriate variables for data-mining tasks.[26]

- *Is the provider willing to provide a custom contract?* Contract negotiations are a critical step in initiating a successful data-mining program. The contract should be as precise as possible and should abstain from nebulous clauses discussing partnership.[27] Moreover, the vendor's standard

contract should not be used, because the standard contract does not customarily include specific performance standards or penalty clauses if the vendor falls short of requirements. Worse, payment schedules in standard contracts may favor the vendor.[28] A custom contract should be written to include service-level measures and a termination clause. The buyer should be particularly suspicious of so-called change-of-character clauses, which state that the buyer may have to pay for any changes in "functionality" throughout the life of the contract. Change-of-character clauses have caused many disputes because of the ambiguous nature of the term "functionality."[29]

Guideline #2: *Build segmentation and predictive models.* Building appropriate segmentation and predictive models necessitates an extensive knowledge of the hotel business. Exhibit 2 provides examples of some of the many ways that hotel guests can be segmented.[30] Transient hotels, convention hotels, extended-stay hotels, and resort properties all segment guests differently. Furthermore, guest segmentation is distinctive for most hotel properties. Hilton's and Marriott's property-management systems segment and code markets at the property level, for instance, since each location has its own particular segments. A given property may serve a set of corporate clients, a group of government clients, and social clients (e.g., weddings and reunions). The segment categories contained in Exhibit 2 can be strung into a large set of combinations. Furthermore, a guest could potentially fit into several categories, which poses a challenge for current data-mining techniques.[31] As a consequence, finding a provider that has experience creating models in the hotel industry is a major benefit. Additionally, even if the provider has hotel experience, it is critical that IT and marketing man-

---

[26] M. Shaw *et al., op. cit.*

[27] P. Lacity and R. Hirschheim, *Beyond the Information Outsourcing Bandwagon* (New York: John Wiley & Sons, 1995).

[28] *Ibid.*

[29] *Ibid.*

[30] See: P. Kotler, J. Bowen, and J. Makens, *Marketing for Hospitality and Tourism*, second edition (Upper Saddle River, NJ: Prentice-Hall, 1999).

[31] M. Shaw *et al., op. cit.*

agers work closely with the provider to segment the market and build predictive data-mining models.

Once a data-mining model is built, confirmatory testing must be conducted to assess its predictive accuracy. For instance, a model designed to predict who will respond to a promotion should be based on a prior offering in which it is known who did or did not respond. After the model is constructed, a "holdout" group from a previous promotion can be analyzed to verify reliability. If the holdout predictions do not replicate the results of the past promotion, then the model may not be significantly predictive. To further enhance accuracy, a score can be assigned to the model based on the level of agreement between the holdout group and the entire group. Subsequent refined models can then be tested and scored. Another standard approach to model validation involves drawing two random samples from the data. The first sample is used as a calibration sample to build the model, while the second is used as a holdout sample to evaluate the model built from the calibration sample.[32] The validation process requires a knowledgeable IT professional, because when data subtleties that arise only in the sample are used to build the model, the model may be highly predictive of the sample but biased with regard to the population.[33] This is called overfitting the data. To avoid creating a biased model, the IT professional must be knowledgeable of the analytical procedure and possess a basic understanding of the hotel segment and promotional scenario from which the sample was extracted.

Guideline #3: *Collect data to support the models.* Accurate data collection is critical for successful data mining. The major obstacle to effective data mining, however, is inadequate data gathering and input.[34] Data problems lead to a decrease in the value of any data warehouse, in addition to diminishing the value of proposed models.[35]

Problems with data are related to one or more of at least three different shortcomings.

The first possible difficulty involves missing or inaccurate data. For example, when occupation information is available for only 15 percent of a data set, it is difficult to create a profile of customer occupations. Then again, it's a problem if the data file contains occupation information for 90 percent of the population, but the accuracy of the information is

> **Finding a data-base expert who has experience creating models in the hotel industry is a major benefit.**

poor. Hotel corporations can reduce inaccuracy of this kind by asking guests for their current occupation.

A second obstacle is poorly coded data. Databases must have standards regarding data formats, text case, and redundant codes.[36] Although some software automatically formats the data properly, most do not. Problems then occur when data-input sources are added over an extended time and no one has ensured that the data entering the warehouse is properly formatted. This would occur, for instance, if, when original data-mining technology was installed, predictions were made based on the reservations system and the property-management system, but then a subsequent decision was made to input data from guest-satisfaction surveys. Problems would transpire when additional data inputs are not standard or are coded improperly. For example, some models require continuous and ordinal data, while others demand categorical data fields or binary constructs.[37]

A third potential problem involves using homonyms (that is, putting the same label on two or more different data elements) and synonyms (that is, using two different labels for the same data

---

[32] P.R. Peacock, "Data Mining in Marketing: Part 2," *Marketing Management*, Spring 1998, pp. 15–25.

[33] M. Shaw *et al., op. cit.*

[34] M. Smith, "Refining Raw Data," *Printing Impressions*, Vol. 43, No. 9 (2001), pp. 36–37.

[35] M. Shaw *et al., op. cit.*

[36] L. Stevens, *op. cit.*

[37] Siragusa, *op. cit.*

element).[38] While it may seem tautological to advise precluding this occurrence, the most common culprit is a new user on the system. It is common for hotel and IT professionals to change companies from time to time. Turnover causes coding problems when new employees bring their old labels and fail to use their new employer's framework.

Guideline #4: *Select the appropriate tools for analysis and prediction.* Numerous analytical tools can be employed to transform data into useful information. Some of the less-common analyti-

> # Training is a key to the effective implementation of data-mining systems.

cal tools used by data-mining software include regression models, factor analysis, cluster analysis, structural equation modeling, and self-organizing maps. On the other hand, the most common statistical methods used in data-mining applications are decision trees, neural networks, and genetic algorithms. As previously mentioned, a decision tree is a rule-based model constructed of nodes (decision points) and branches (connections between nodes) that reach numerous outcomes based on traveling through two or more nodes. A neural network is a nonlinear predictive model that resembles a biological neural system and has the ability to learn through training. Last, a genetic algorithm is a learning-based model founded on the concept of evolution. That is, partial solutions to a scenario compete with each other, and then the best solutions are used for further problem solving.[39]

Most of the statistical methods employ techniques that achieve a desired outcome. Likewise, each methodology has strengths and weaknesses,

and each is appropriate for a specific scenario. Therefore, the most effective results emanate from data miners who have the expertise to select the most appropriate statistical method for a given scenario and the hotel's intended goals.[40] For instance, a positive attribute of genetic algorithms is that they converge on an optimal solution, but the method is most applicable to large databases since arriving at a valid outcome may require many generations of competing solutions. Likewise, there are also pros and cons associated with neural networks. They are beneficial in analyzing complex data because of their ability to discover unusual trends, but monitoring accuracy is difficult because many intricate relationships are handled invisibly by the methodology.[41]

Guideline #5: *Demand timely output.* Timeliness is critical in making marketing decisions. The length of time required to produce output varies widely among data-mining packages. Before Hilton Corporation upgraded its data-mining technology, for instance, the reports that managers requested from IT would take three to six weeks to arrive. "By the time they'd get the report, it was often too late to act on it," said Joanne Flinn, vice president of leisure marketing. With the new technology, managers receive reports in 30 minutes or less.[42]

Guideline #6: *Refine the process.* By its nature, data mining involves knowledge that evolves over time. Never complete, data mining involves a continuous cycle of inputs and outputs based on models that must be modified and refined as conditions change in the competitive environment. Flexibility is needed to adapt the established models and processes to changes that occur.[43] Refinement consists of three actions:

(1) Chart progress toward initial goals. Use the forecasting function of data mining to regularly set new goals.

(2) Compare and contrast the characteristics of the clustering output with the attributes of the classification output. When necessary, modify predictive

[38] J. Chopoorian, R. Witherell, O. Khalil, M. Ahmed, "Mind Your Business by Mining Your Data," *S.A.M. Advanced Management Journal,* Vol. 66, No. 2 (2001), pp. 45–51.

[39] J. Hair, R. Anderson, R. Tatham, and W. Black, *Multivariate Data Analysis*, Fifth Edition (Upper Saddle River, NJ: Prentice-Hall, 1998).

[40] Siragusa, *op.cit.*

[41] J. Hair *et al., op. cit.*

[42] Stevens, *op. cit.*

[43] R. Cline, *op. cit.*

models based on changes in the size or structure of customers' market segments. For instance, notable variances in purchase patterns in one segment and similarities in purchase patterns among other segments may lead to refinement of the segment, usually by adding a new criterion or dimension. This action can result in the discovery of previously undetected segments with homogenous characteristics.

(3) In addition to clustering and classification features of data mining, also use the associations and deviation-detection functions to assess the effects of promotions. Maintain a promotional history table in the database to use as a learning tool for future campaigns and models.[44]

Guideline #7: *Hire a well-trained staff and a knowledgeable IT manager.* Information technology was initially viewed by the hotel industry as a back-office function that supports the finance and accounting areas.[45] The industry has advanced far beyond this view during the past decade. In two sessions sponsored by the International Hotel and Restaurant Association (IH&RA), one in Singapore in 1997 and the second in Nice, France, in 1998, hotel-industry leaders pondered the role of technology. Among the conclusions reached were: "Going forward, technology will be the most competitive weapon for any hospitality company. If hospitality organizations want to compete successfully, they must do so by using technology to drive value to both the customer and to the firm."[46] However, implementing such recommendations at the property level can be a difficult task.

Training is a key to effective implementation of data-mining systems. Productive data mining requires two-fold proficiency among both IT managers and those who interpret the outputs.

The hotel's IT managers must also be proficient with the data-mining system because the system requires continuous refinement. Just as

market segments, sources of data, and property goals change, so must predictive models and analyses be modified and refined. It is an unsound policy for the IT staff to be totally dependent on the provider's recommendations for refinement and alterations. Instead, the IT staff and data-mining provider should work together with their common goal being to maximize the technology's effectiveness. The most-effective data-mining projects occur when IT managers and providers collaborate and share project information.

> **Data-mining technology should be used in conjunction with other forecasting and research techniques.**

Second, adequate training must be provided to all potential users of data-mining outputs. At the corporate level this includes the marketing staff, operations managers, and those developing new properties. Users at the property level include general managers, directors of sales and marketing, and the sales staff. Users must be instructed about the available reports and how to properly interpret the information. Since the information is used for decision making, it is important for users to understand the boundaries and limitations of the information.

Boundaries and Limitations

Technology must serve managers' purposes, rather than dictate processes.[47] Along that line, data mining cannot capture all the information relating to what drives consumer behavior. Data mining is simply one of a number of research methods that help predict travelers' demand trends. Therefore, data-mining technology should be used in conjunction with other forecasting and research techniques. With this in mind, managers should be aware of the following four limitations of data-mining technology.

---

[44] Siragusa, *op. cit.*

[45] Cline, *op. cit.*

[46] M. Olsen and D. Connolly, "Antecedents of Technological Change in the Hospitality Industry," *Tourism Analysis,* Vol. 4 (1999), p. 29.

[47] C. Chudnow, "Knowledge-management Tools," *Computer Technology Review,* Vol. 21, No. 11 (2001), pp. 28–29.

Psychological determinants of demand

Education	Relaxation
Escape	Self-discovery
Family bonding	Sexual opportunity
Prestige	Social interaction

Source: P. Kotler, J. Bowen, and J. Makens, *Marketing for Hospitality and Tourism*, second edition (Upper Saddle River, NJ: Prentice-Hall, 1999).

Limitation #1: *Data mining analyzes only data collected from existing customers.* Data-mining software generates information by analyzing data patterns derived from the company's reservation, property-management, and guest-loyalty-program systems. Patterns thus detected can help predict the actions of current guests in the system and of those with similar needs and wants. Data-mining technology does not, however, provide information about market segments not found in the company's databases. Moreover, a market segment that is currently small but is on the verge of experiencing substantial growth may not be detected by data mining.

Another blind spot is the data in competitors' reservation systems. A key question in planning a marketing strategy in the hotel industry is: Who are my competitors' guests and where are they coming from? Data-mining technology is unable to answer those questions.

Limitation #2: *Databases used in the mining process are often hotel-brand specific.* Just as data mining cannot analyze competitors' markets, it also creates prediction models that are brand specific. Thus, corporations that operate multiple brands often must create a data warehouse and conduct data mining for each brand. This is also true for the franchisees that may have a portfolio comprising, say, six Holiday Inns and four Marriotts.

Brand-specific marketing information is useful for the brand's corporate office to plan marketing programs, which is largely what franchisees purchase. Conversely, brand-specific marketing information may not be helpful if the hotel corporation that franchises numerous brands wants to predict customer demand based on a multiple-brand portfolio.

Limitation #3: *Data mining may not segment travelers by psychographic traits.* Segmenting consumers based on psychographic traits, such as personality and lifestyle, can be useful in the hotel industry. This is because psychology and emotion play significant roles in the hotel guest's decision process. That is, as seen in Exhibit 3, a traveler may select a destination for a variety of psychological reasons.[48] One limitation of data mining is that common system inputs do not account for psychological factors that influence a traveler's purchase decision.

A time-tested tool used in understanding hospitality demand trends is Stanley Plog's psychographic scale.[49] Many key drivers of demand identified by Plog, such as personality distribution among travelers (e.g., dependables, venturers, and centrics), are not common inputs into data-mining systems. Hotels can acquire this information from customer surveys.

Limitation #4: *Data mining does not provide information about consumers' thought processes.* It is important to engage consumers in research to better understand their thinking. Information generated by data mining does not account for the fact that approximately 80 percent of human communication is nonverbal.[50] Interviews and focus groups are both useful methods for gathering information about the needs and wants of hotel guests. The insight gained from those techniques is difficult to capture in the statistical data-mining outputs. That is why it is important to step back and ask what the hotel guest's inherent needs are and what the product is really about. This involves conducting in-depth conversations with guests. At times, improved insight and perspective are gained from talking with three customers for two hours rather than by surveying a thousand customers.[51]

[48] Kotler *et al., op cit.*

[49] S. Plog, "Why Destination Areas Rise and Fall in Popularity," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 42, No. 3 (June 2001), pp. 13–24.

[50] G. Zaltman, "Rethinking Market Research: Putting People Back In," *Journal of Marketing Research*, Vol. 34, No. 4 (1997), pp. 424–437.

[51] K. Ohmae, *The Borderless World* (New York: McKinsey & Company, Inc., 1999).

Conclusion and Managerial Implications

Data-mining technology can be a useful tool for hotel corporations that want to understand and predict guest behavior. Based on information derived from data mining, hotels can make well-informed marketing decisions—including who should be contacted, to whom to offer incentives (or not), and what type of relationship to establish.

Data mining is currently used by a number of industries, including hotels, restaurants, auto manufacturers, movie-rental chains, and coffee purveyors. Firms adopt data mining to understand the data captured by scanner terminals, customer-survey responses, reservation records, and property-management transactions. This information can be melded into a single data set that is mined for nuggets of information by data-mining experts who are familiar with the hotel industry.

However, data mining is no guarantee of marketing success. Hotels must first ensure that existing data are managed—and that requires investments in hardware and software systems, data-mining programs, communications equipment, and skilled personnel. Affiliated properties must also understand that data mining can increase business and profits for the entire company and should not be viewed as a threat to one location. As seen in the Harrah's example, implementing a data-mining system is a complex and time-consuming process.

We advise hospitality managers to adopt a data-mining system and strategy if they have not done so. Guidelines presented in this paper—including how to select and manage the data-mining provider—offer guidance for implementing a viable data-mining strategy. Since data mining is in its initial stages in the hotel industry, early adopters may be able to secure a faster return on investment than will property managers who lag in their decisions. Hotel corporations must also share data among properties and divisions to gain a richer and broader knowledge of the current customer base. Management must ensure that hotel employees use the data-management system to interact with customers even though it is more time consuming than a transactional approach. ■

Vincent P. Magnini (photo at far left) is a Ph.D. candidate at Old Dominion University (vmagnini@odu.edu). **Earl D. Honeycutt, Jr.,** Ph.D. (middle photo), is a professor at Love School of Business at Elon University (ehoneycutt@elon.edu), where **Sharon K. Hodge,** Ph.D. (photo at immediate left), is an assistant professor (shodge@elon.edu).