December 2006

Using Data Mining to Facilitate More Efficient Resource Allocation and Training for IT Support in Large Organizations

Charles McCann
Oklahoma State University

Marilyn Kletke
Oklahoma State University

Follow this and additional works at: http://aisel.aisnet.org/amcis2006

Recommended Citation
http://aisel.aisnet.org/amcis2006/186

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2006 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Using Data Mining to Facilitate More Efficient Resource Allocation and Training for IT Support in Large Organizations

Charles L. McCann  
Oklahoma State University  
Charles.Mccann@okstate.edu

Marilyn G. Kletke  
Oklahoma State University  
Marilyn.Kletke@okstate.edu

ABSTRACT

The information technology (IT) support function in business organizations has evolved from “end user computing” support in the 1980’s to what is now referred to as “information centers” or “IT Help Desks.” In recent years the IT Help Desk has become a strategic focus of management. In a top-down performance-enhancing effort, management often employs statistics to help make system resource allocation and training decisions. We propose that data mining of and nugget discovery in customer call data and associated customer satisfaction surveys can be used to approach the training issue from the bottom up, thereby making more efficient use of resources. This is illustrated through an empirical study from a University IT Help Desk over a period of several months. Results show that data mining can lead to specific, focused bottom-up policy decisions that are more resource-efficient than the traditional broadly applied training programs applied top-down across the system.

Keywords

Data Mining, Nugget Discovery, IT Help Desk, Information Center, End-user Computing, Customer Satisfaction, Success Factors, Training Efficiency, Resource Allocation

INTRODUCTION

The information technology (IT) support function in business organizations, represented by different names across the years, has been the subject of extensive research in the MIS literature since the early 1980’s. The goal of the IT support function has always been to support what used to be called “end user computing,” and is now often referred to as “end users” or “users” or “clients.” Such support occurred first in a mainframe environment, and today occurs in a distributed computing environment as well. The terms “information center” and “help desk” evolved over time to represent the organizational entity charged with “end-user support.” Initially the information center (IC) was intended solely to help users develop and use software, but in recent years it has often taken on a strategic role for the organization (McLean, Kappelman, and Thompson, 1993). Performance of the IT Help Desk is the subject of much research in today’s organizational environment.

Management efforts to improve the performance of the IT Help Desk in its newer strategic role have focused on increasing customer satisfaction, a critical success factor (CSF) that originated in the key end user variable “user satisfaction” in the taxonomy suggested by DeLone and McLean (1992). Many organizations are now collecting large amounts of survey data from clients by requesting a post-service evaluation every time a client uses the IC services. Performance metrics for IT Help Desk operations that have been found to directly impact customer satisfaction are “percentage of calls closed on first contact” and “average abandonment” (Feinberg, Kim and Hokama, 2000). These metrics can be found in responses obtained on the client surveys. In general, management sets and implements policies and procedures designed to improve the numbers on these surveys. One approach to improving the performance of an IC is for management to make a general decision that impacts the system as a whole, or the policies and procedures of the IC. For example, to improve the CSF client satisfaction, one procedure that management implements is training of IT Help Desk technical support personnel. Typically organizations have incrementally improved the performance of ICs by implementing training efforts directed at the entire group of first-level technical support personnel of the IC (e.g., Webster and Kral, 2002). Ward (2005), too, reports on creating training tools for all IT Help Desk technical support personnel. The philosophy is that if all first-level personnel go through the same extensive training, then they will better serve the clients, who will express higher satisfaction with the service, which will thus reflect an improved performance of the IC. This approach to training is a top-down system approach in that it implements the same training across all first-level personnel in the system; the benefits are expected to ripple through the
system and result in general improvement of the IC performance. Additionally the training is focused on the overall system, not a specific portion of the system.

The costs associated with system-level training are significant, and an organization can benefit by finding ways to reduce these costs without reducing benefits. Webster and Kral (2002), for example, underscore the need in their study where budgetary considerations were paramount and projected budget cutbacks forced a reduction in the scope of the training program. It is difficult to estimate the impact on CSF’s of reducing a general training program for IT Help Desk technical support personnel.

An alternate approach is to try to improve a given CSF at the lowest level in the system. For example, instead of applying process improvement or training across the board to all first-level technical support personnel, one could instead examine specific scenarios that yield a pattern of less than satisfactory performance. The process improvement or training would be prescribed and directed at a specific type of problem, or sub-problem, or even one particular technical support individual, as opposed to the entire system.

To this end, we propose that levying data mining techniques against large amounts of available “service ticket” data and associated customer satisfaction surveys can provide information, for the organization that collects it, that will make it possible for management to make better decisions about allocating resources. In the empirical example illustrated herein, the information yielded by a data mining technique can be used by management to make and apply specific decisions to a subset of the system that will improve IC performance as well or better than providing identical aggregate resource commitments focused on the entire system. We argue that required resources will be significantly less than those consumed by the overall approach. The contribution of this manuscript is the methodology of isolating areas of weakness that management can address directly.

OBJECTIVE
The objective of this paper is threefold:

1. to gather variables known as CSF’s that have been identified over the years as being principal components of IT Help Desk success; and to identify performance metrics associated with these CSF’s
2. to identify standard performance metrics that are typically used to measure these CSF’s
3. to illustrate through an empirical example how a data mining approach can help improve resource allocation and training to improve IT Help Desk performance

BACKGROUND
In the 1970’s end user computing (EUC) was a phenomenon that was associated with mainframe computing. With the advent of the personal computer, the nature of EUC began to change. In their seminal article, Rockart and Flannery (1983) provided a discussion of EUC and how management could successfully address the changes. Hammond (1982), in his visionary manuscript, defined and constructed the details of the organizational entity described as an IC to support EUC that became the basis for research in the next decade. In tandem with the IC idea introduced by Hammond (1982), many researchers addressed EUC and associated issues. EUC was listed as a key issue from the management perspective (Brancheau and Wetherbe, 1987; Dickson, Leitheiser, Wetherbe, and Nechis, 1984).

During the late 1980’s and the 1990’s, EUC was gradually folded into the overarching issue of organizational computing (Watson, R., Kelly, G., Galliers, R. and Brancheau, J., 1997), and with it came the rise of issues associated with ICs, call centers, and IT Help Desks. Vijayaraman and Ramakrishna (1990) conducted an analysis of successful versus unsuccessful ICs, and end user satisfaction was identified as the key success factor. Sometime later IT Help Desks were categorized as organizational subsets of ICs (Pitt, Watson, and Kavan, 1995), and the measuring of user satisfaction was accomplished relative to the IT Help Desk. Wilson (1997) underscored the growing respectability and, in fact, essential nature of the IT Help Desk.

CRITICAL SUCCESS FACTORS AND PERFORMANCE METRICS
As research developed, studies focused on CSF’s that were associated with success of the ICs and IT Help Desks. As accord grew on CSF’s, research began to relate CSF’s to variables that could be part of performance metrics. Consistently through time the overarching CSF has been and continues to be user satisfaction. Associated with CSF’s are various performance variables that can be measured to explain or predict the CSF. Many studies highlight variables that contribute to user satisfaction, either directly or indirectly. Magal, Carr and Watson (1988) identified 26 CSF’s from the literature that were applicable to ICs and that fit into three major categories for user satisfaction and IC success: a competent staff,
communication with end users, and top-management support for the IC. Vijayaraman and Ramakrishna (1990) classified IC support services with regard to adaptability, timeliness, availability, dependability and adequacy; and an EUC-related CSF was good user training and education. Magal (1991) focused on performance metrics associated with user satisfaction, and classified 16 of them into three categories: quality of IC services, quality of user-developed applications (UDA), and quality of user self-sufficiency-related items. Pitt, et al. (1995) attributed success to physical facilities, and four variables related to technical support personnel: reliability, responsiveness (willingness to help and speed), assurance (trust and confidence issues), and empathy (caring attention). Essex, Magal, and Masteller (2001) determined that the main contributors to the CSF user satisfaction were quality of UDA’s, user self-sufficiency, quality of individual staff, quality of services, and facilitation of end-user computing. Feinberg, Kim, and Hokama (2000) found these variables that impact customer satisfaction in business organizational call centers: average speed of answer, queue time, first call resolution, abandonment rate, average talk time, adherence (technical support personnel there as advertised), average work time after call, % of calls blocked, time before abandonment, inbound calls per technician in 8-hour shift, technician turnover, total calls, and service levels. Shaw, Delone, and Niederman (2002) found these factors to support user satisfaction across all user groups: IS staff response time, new software upgrades, positive attitude of IS staff, and data security and privacy. Factors that were significant for a single user group were documentation to support training, IS staff technical competence, and ease of access to computing facilities.

Some studies have examined user dissatisfaction. Heckman and Guskey (1998), for example, classified contributors to user dissatisfaction into service delivery failures (help unavailable, too many helpers, problem unresolved), customer situation before call (on deadline, customer error, novice, previous experience), and IT technician characteristics/behavior (competence, attitude, effective communication, speed of service, extraordinary behavior).

These studies typify an area of research that has set the groundwork for and developed some specific call performance metrics to evaluate user satisfaction. These metrics can be classified into these areas: quality of staff/service, time to solution, quality of UDAs (if applicable), and how the resolution of the problem contributed to the self-sufficiency of the customer. Two major performance metrics are first call resolution (a dichotomous variable) and abandonment rate (the percent of callers who hang up before the problem is resolved).

**IT HELP DESK IMPROVEMENT**

Management wants to improve the performance of IT Help Desks. They judge improvement by performance variables such as are identified above. It isn’t clear in the call center literature which variables are related to caller satisfaction (Feinberg et al., 2000). In making decisions to improve performance there have been two main approaches to decision making: statistical analysis and on-line analytical processing (OLAP) with drill-down analysis. We introduce a third: data mining; and we discuss how data mining, specifically classification and nugget discovery, can be used in conjunction with the other methods to gain greater insight into the problem domain. Our focus in the data herein is on classifying current behavior rather than on predicting future outcomes; the objective is that management use the resulting information to make policy decisions that will improve future outcomes.

**Statistical Measures**

An example of using statistics occurs in the top-down way management often makes training resource allocation decisions. They examine one variable in their data; for example, first call resolution. They then make the decision to increase the general level of first call resolution because the literature has shown that it has a huge impact on caller satisfaction. So they decide, for example, to train first-level IT Help Desk personnel on the software that is the most frequent subject of calls; this can be found by simple statistical analysis of the data. This training, required of every first-level technician, is supposed to engender benefits that will ripple through the system; more callers will be more satisfied, and the overall satisfaction for the IT Help Desk should marginally increase.

**OLAP**

Using OLAP on the data can improve on the training decision that is made using simple statistics. Management drills down from the general data, to specific dimensions or classes of problems, and looks at the first call resolution percentage for each. Then management selects a subclass with a low first call resolution percentage that they can impact by, for example, training first level technicians on one specific issue. Management has a hierarchical display of first call resolution, with a subclass of a particular type of problem being the piece of information obtained by OLAP. System and training benefits should increase the first call resolution percentage by a significant amount for the one subclass they addressed; and the overall satisfaction for the IT Help Desk should increase.
Data Mining

We propose augmenting the above methods with an approach to IT Help Desk improvement aided by data mining. According to Fayyad and Uthurusamy (2002) data mining is the process of automated discovery of interesting patterns, trends, and correlations hidden in databases by sifting through large amounts of data. Data mining was first recognized as a formal information technology in the mid-1990s (Leavitt 2002). Classification is a specific data mining strategy that is applicable to this study. Existing classification methods for knowledge extraction either seek to describe current behavior by locating patterns in the data; or they seek to find patterns that will result in building predictive models for the entire dataset. A related data mining application is called partial classification, or nugget discovery. This is an application of classification in which patterns or anomalies can be related to a pre-determined class or classes (Abbass, Sarker, and Newton, 2002), and simple and/or complex rules can be generated to describe the data. It should be noted here that classification and nugget discovery have different objectives. Where complete classification is frequently evaluated on overall accuracy in classifying new data, this metric is not applicable to nugget discovery because the large and often complex structures produced in classification often do not reveal the interesting nuggets that may be found in small parts of the dataset (Quinlan, 1987).

METHODOLOGY

The Data

To illustrate this concept, we selected data from an IT Help Desk at a large, south-central research university. Each call to the IT Help Desk initiates a “service ticket,” which is used to track the incident from then until closure. Over time a large dataset can be collected. This IT Help Desk utilizes approximately 30 first-level technical support persons connected to a Knowledge Base of problem solutions. If the first-level technician cannot resolve the issue, the ticket is sent to a second-level support specialist; it will then go to a third-level support if it is not resolved. On closure the client is requested to complete an on-line survey reflecting his/her experience (see Appendix). The call data together with the survey data forms the dataset used for this study.

The Survey

The survey instrument is structured to reflect the major variables discussed above that impact user satisfaction. Questions are set up using a 5-item Likert rating scale. The user is asked first to rate his or her satisfaction with the overall quality of service, and this is used as the CSF in this analysis. Other questions that identify impacting variables include quality of staff, competence of staff, and staff response time. Three additional questions asked for an indication of satisfaction with the IT Help Desk hours of operation, an indication of service expectation for next time, and an indication of willingness to recommend the IT Help Desk service to others. Variables tracked for each call from the IT Help Desk side include IT technician name/id, resolution, category and subclass of the problem, time to solution, and some history of the resolution process. The average response rate on customer surveys is between 12% and 15%. The data set used in this study contains approximately 31,000 service tickets from over a year of data. Of these 31,000 records 2,300 were submitted via the web or through email and were excluded from investigation. Of the remaining 28,699 records remaining, 2,146 satisfaction surveys were returned, for approximately a 7.5% response rate.

A data mining approach applied to this dataset can provide information that will direct system improvement and training to specific focal areas. The classification method, in the form of nugget discovery for the example included in this study, will provide a set of rules that characterize the types of issues that lead to an inability to resolve the issues on first contact. We will then be able to use these patterns to highlight the need for management attention in the identified areas. Note that we are not interested in prediction or in an accurate description of negative examples; we are interested in a nugget whose consequence is user dissatisfaction.

The Approach

As a specific example of how this multi-method approach can be useful, we applied it to the service ticket and associated user satisfaction data described above. Table 1 shows the variables that are in the dataset; they are for the most part categorical. The Likert rating scale in the table includes Extremely Satisfied, Satisfied, Neutral, Dissatisfied, and Extremely Dissatisfied. We are interested in guiding management to make decisions that will reduce customer dissatisfaction, thereby increasing customer satisfaction. First, from overall service quality was selected as the dependent variable representing customer satisfaction. We then used multiple regression analysis to determine which factors had the highest impacts on overall customer satisfaction. We found that “Q2 – Ability to Solve” and “Q5 – Timeliness of Resolution” were by far the most significant factors driving overall customer satisfaction (significant at the .001 level). Given these findings along with previous research findings it is apparent that first call resolution, which is highly correlated with these factors, contributes
significantly to overall customer satisfaction. This dichotomous variable is an item of interest for the example of our data mining approach.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CallerID</td>
<td>Client on ticket for this trouble call</td>
<td>6-digit ID</td>
</tr>
<tr>
<td>OpenID</td>
<td>Technician who first took the call</td>
<td>6-digit ID</td>
</tr>
<tr>
<td>CloseID</td>
<td>Technician who solved the problem</td>
<td>6-digit ID</td>
</tr>
<tr>
<td>FirstCall</td>
<td>Was the problem resolved on the first call?</td>
<td>True/false</td>
</tr>
<tr>
<td>Category</td>
<td>Problem category</td>
<td>String</td>
</tr>
<tr>
<td>Class</td>
<td>Class of problem within category</td>
<td>String</td>
</tr>
<tr>
<td>Product</td>
<td>Software/hardware system with which the problem occurred</td>
<td>String</td>
</tr>
<tr>
<td>Q1</td>
<td>Overall service quality</td>
<td>Likert scale 1-5</td>
</tr>
<tr>
<td>Q2</td>
<td>Technician’s ability to solve</td>
<td>Likert scale 1-5</td>
</tr>
<tr>
<td>Q3</td>
<td>Technician’s ability to communicate information about the service performed</td>
<td>Likert scale 1-5</td>
</tr>
<tr>
<td>Q4</td>
<td>Responsiveness of the IT Help Desk to the problem</td>
<td>Likert scale 1-5</td>
</tr>
<tr>
<td>Q5</td>
<td>Timeliness of resolution of the problem</td>
<td>Likert scale 1-5</td>
</tr>
<tr>
<td>Q6</td>
<td>Technician’s courtesy</td>
<td>Likert scale 1-5</td>
</tr>
<tr>
<td>Q7</td>
<td>Overall quality of experience with technician</td>
<td>Likert scale 1-5</td>
</tr>
</tbody>
</table>

Table 1. Variables Associated with IT Help Desk Service Tickets and Associated Customer Satisfaction Surveys

We next applied targeted data mining to provide useful information for IT Help Desk services. Consider the application of one common modeling structure: decision trees. In this method, available variables are examined with respect to a given dependent variable to find the one independent variable that contributes the most information about the dependent variable. This one independent variable is selected as the first node in the decision tree, and the individual cases are then split according to which subset of the node they belong. This decision variable becomes a parent node with child nodes representing the subsets of the decision variable. The algorithm steps along adding nodes to the decision tree and children to the nodes until leaf nodes at the ends of the paths are either pure, contain a desired minimum number of cases, or fail to add to the information in the tree. Selection of nodes contributing the most information leads to the generation of a broader, shallower tree than selecting other decision nodes.

In general, decision trees can be used in two different ways. The first is to build the model to learn the concepts and classify the behavior of an existing data set. In this case the goal is descriptive, with the objective to find patterns related to some class or classes. The second is to build the model to predict outcomes in new data by looking at the presence and levels of the model’s decision nodes in future (other) situations. In our example we wish to use the decision tree results to recommend training decisions based on current behavior; so we are using the decision tree in the first way. Furthermore, we are searching for an important pattern that is found in a relatively small portion of the entire dataset. This task, nugget discovery, will lead to information that will guide management policy decisions with respect to modifying the performance of the IT Help Desk.

RESULTS

From the result of the multiple regression analysis referred to above, we selected “resolved at first contact” as our (dichotomous) variable that would serve as the root node for our decision tree. We applied the decision tree methodology to classify our ticket data by “resolved at first contact” and “not resolved at first contact.” This human intervention allows the decision tree algorithm to work within the portion of the dataset that may contain the nugget we are seeking. From the resulting decision tree we observed the following: on all tickets 89% (25,545) are resolved at first contact where as 11%
(3,153) are not. This finding illustrates the typical situation that management finds themselves in: only a relatively small percentage of the tickets (11% for our situation) are not resolved at first contact; but improving first contact resolution offers the best opportunity for improving customer satisfaction. Management cannot apply resources efficiently with this shallow level of understanding. Hence, nugget discovery, where our “nugget” is represented by the “not resolved at first contact” branch of the decision tree, becomes useful. Looking down this branch to the next node in the decision tree, we determined that tickets which are related to printing (2,194) only have a 50% (1,110) first call resolution rate. Stated differently, one-third of the issues which are not resolved at first contact are printer-related; moreover, 55% (606) of those issues are “SCT Printer”-related issues. Management can use this nugget to investigate the cause for the relatively low first call resolution rate. This represents a focused and targeted approach that will be much more cost effective than any general process improvement or training program that they might decide to implement. If management could eliminate these 606 failures, the overall first contact resolution rate would increase from 89% to 91%, thereby increasing customer satisfaction. This nominal 2% increase in first contact resolution rate takes on added impact when it is further seen that this is approximately 20% of the 11% of all calls that were not resolved at first contact. Additional improvements can be made by continuing to traverse the tree; however, each step further from the root node yields a lower percentage improvement.

SUMMARY AND CONCLUSIONS

This multi-method approach highlights the importance of matching the appropriate data analysis technique with the data available. Additionally it shows that data mining is a powerful tool which can be used in an IT Help Desk environment to highlight in a focused manner specific sources of customer dissatisfaction. Specifically, nugget discovery, a form of classification, allows better description of factors that can be impacted by management decisions. In short, management can use the extracted information to make focused resource allocation and training decisions that are more efficient than general system-based decisions. As an example, it has been shown that decision tree analysis, specifically used in nugget discovery, can be used to direct management in their efforts to increase customer satisfaction. Specifically, through the use of data mining, management can selectively improve processes or train individual IT Help Desk technicians on specific products/areas as needed. The dollars spent in addressing the problems are focused and smaller than they would be in implementing broad process improvement or training programs.

FUTURE RESEARCH

More research is needed to explore the potential of targeted data mining techniques in guiding IT Help Desk management to more efficient resource allocation and focused training directives for IT support.

First, research is needed on exactly what variables impact customer satisfaction in IT Help Desk environments. Feinberg et al. (2000) indicate that solid research studies linking variables to customer satisfaction in call centers is non-existent. Identification of these variables will facilitate additional research on the benefits of data mining to provide information directly related to improvement of IT Help Desks. Other CSF’s associated with IT Help Desk performance need to be identified and associated with causal variables. Performance metrics should be identified for these CSF’s.

Additionally, future research should look at patterns that data mining can reveal for different sets of CSF’s of IT Help Desk performance. The approach discussed above can be extrapolated to apply to any subset of CSF’s. Once the variables that impact the CSF’s are identified, then data mining can be used to produce information that will impact management decisions regarding the IT Help Desk. The use of nugget discovery, along with measures of interest, can be developed and documented for various CSF’s. From patterns that are found management can implement focused policy changes. Resource expenditures will be less than they would in broad-based statistically-based actions, and they should result in improvement to the performance metrics for the particular CSF(s), and thus to the IC.

Last, research from a knowledge management perspective will contribute additional valuable insights to this problem. A new focus on IT Help Desk implementation methodologies has been introduced by Gonzalez, Giachetti, and Ramirez (2004), who suggested a knowledge management system for an IT Help Desk. They analyzed the performance of a simulated IT Help Desk that used such a system, and demonstrated that the time to solution to all but the system-critical calls was significantly reduced. Further research along this path will help build knowledge management systems for IT Help Desk management and will yield valuable results to practitioners needing to improve IT Help Desk performance.
APPENDIX

IT Help Desk Survey Facsimile

Q1. How satisfied were you with the overall quality of our service?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Q2. How satisfied were you with the technician’s ability to solve your issue?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Q3. How satisfied were you with the technician’s ability to communicate information about the service performed?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Q4. How satisfied were you with the responsiveness of the Help Desk to your request?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Q5. How satisfied were you with the timeliness of your resolution?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Q6. How satisfied were you with the technician’s courtesy?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Q7. How satisfied were you with the overall experience with the technician?
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

I am ______________ with the current operational hours of the Help Desk.
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

I believe that if I contact the Help Desk again, I will be ______________.
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

I would be ______________ in recommending the Help Desk’s service to others.
   - Very Satisfied
   - Satisfied
   - Neutral
   - Dissatisfied
   - Very Dissatisfied

Please put any comments or any additional information here:
REFERENCES