1-1-2015

Enabling Self-Service BI Through a Dimensional Model Management Warehouse

Karen Corral
Boise State University

David Schuff
Temple University

Gregory Schymik
Grand Valley State University

Robert St. Louis
Arizona State University

Enabling Self-Service BI through a Dimensional Model Management Warehouse

Full Paper

Karen Corral
Boise State University
karencorral@boisestate.edu

David Schuff
Temple University
schuff@temple.edu

Greg Schymik
Grand Valley State University
schymikg@gvsu.edu

Robert St. Louis
Arizona State University
stlouis@asu.edu

Abstract

The promise of Self-Service Business Intelligence (BI) is its ability to give business users access to selection, analysis, and reporting tools without requiring intervention from IT. However, while some progress has been made through tools such as SAS Enterprise Miner, IBM SPSS Modeler, and RapidMiner, analytical modeling remains firmly in the domain of IT departments and data scientists. The development of tools that mitigate the need for modeling expertise remains the "missing link" in self-service BI, but prior attempts at developing modeling languages for nontechnical audiences have gone largely unadopted. This paper seeks to address this unmet need, bringing model-building to a mainstream business audience by introducing a structured methodology for model formulation specifically designed for practitioners. We also describe the design for a dimensional Model Management Warehouse that supports our methodology and demonstrate its viability using an illustrative example. The paper concludes by outlining several areas for future research.

Keywords

Business intelligence, model management, analytics, modeling, self-service

Introduction

In 1987, Box and Draper wrote: "Essentially, all models are wrong, but some are useful" (p. 424). They went on to say: "Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration." Box and Draper's words are very relevant for today's business intelligence practitioners. The science and art of business intelligence has typically required a team with diverse skills ranging from data storage and retrieval, to model formulation and selection, to the presentation of actionable results to business managers.

Through products such as SAS Enterprise Miner, IBM SPSS Modeler, and RapidMiner, we are seeing the emergence of visual analytics model-building tools in the same way that we saw the emergence of visual programming tools 20 years ago. These tools seek to "democratize" analytics (see Henschen 2014; HBR Analytics Services 2012) through the realization of "self-service" BI, making advanced data analysis accessible to a wider audience. Self-service BI seeks to give business users access to selection, analysis, and reporting tools without requiring intervention from IT. However, just as visual programming tools don't make people better programmers, visual modeling tools don't make people better modelers. In fact, it can make things worse by misleading users into thinking they are doing "good" analytics simply because they are able to complete an analysis. In order to truly democratize analytics, we need tools that support decision-making around the model-building process and not simply mask the complexity of statistics and coding.

Information systems professionals have had a great deal of experience with managing, organizing, and presenting data in both structured (e.g., spreadsheets and databases) and unstructured (e.g., textual

documents) forms. However, model building historically has fallen within the domain of management science (Geoffrion 1987; Kottemann and Dolk 1992; Lin et al. 2000). That must change with the widespread adoption of business intelligence and analytics. For analytics to move beyond the purview of data scientists, business-facing practitioners must employ methodologies and tools that help them: 1) understand the difference between data, documents, and models, and the implications of those differences for model building and management; 2) identify relevant variables and their relationships; 3) assess the usefulness of models; and 4) know when to terminate the model building process.

This paper describes a structured methodology for model formulation specifically designed for practitioners, and the design for a dimensional data store that supports that methodology. We begin by reviewing the literature on data and document retrieval and extend this work to the retrieval of analytical models. We then review the work that was done by management scientists on model management and explain why that work was never sufficiently implemented in practice. To address those shortcomings, we present our approach and discuss why our methodology and underlying data store is uniquely poised to democratize the use of analytics while encouraging "good modeling behavior." We conclude with future directions, describing a research agenda to further develop and test our approach.

Data versus Documents versus Models

Blair (2002), through an analysis of the differences between data retrieval and document retrieval, proposed that the information search process changes based on the type of artifact being targeted (see the first two columns of Table 1). He argued that the task of finding information contained in documents is fundamentally different and more complex than the task of finding data. A data retrieval task is closed-ended and direct with an unambiguous answer – for example, "what grade did Chen receive for the Database Systems course in the fall semester of 2014?" Data retrieval success is characterized by a "correct" (and verifiable) answer. The time it takes to return the answer is dependent only on the speed of the software and hardware executing the query.

Data Retrieval	**Document Retrieval**	**Model Retrieval**
Direct ("I want to know X")	Indirect ("I want to know about X")	Investigative ("I want to find a model that explains X")
Necessary relation between a formal query and the representation of a satisfactory answer	Probabilistic relation between a formal query and the representation of a satisfactory answer	Satisficing relation between a formal query and the representation of a useful model that recognizes tradeoffs between accuracy and complexity
Criterion of success=correctness	Criterion of success=utility	Criterion of success=improved ability to predict, manipulate, or understand X
Speed dependent on the time of physical access	Speed dependent on the number of logical decisions the searcher must make (include or discard)	Speed dependent on the number of modifications required to obtain a useful model

Table 1. Comparison of data, document and model retrieval (adapted from Blair 2002)

In document retrieval, the underlying questions are more open-ended and indirect, and there may not be a single correct answer – for example, "Which students are most likely to graduate?" The formal query often is phrased in several different ways to gather a set of documents that, together, are likely to provide a sufficient answer to the question. For example, queries might include "student success factors," "graduation rates," and "at-risk students." These searches are likely to return multiple results, as it frequently is the case that more than one document will contain relevant information. Document retrieval

success is based on the utility of the documents returned for formulating an answer to the question being researched. The time it takes to formulate an answer is dependent on both how many documents are returned, and the speed with which the searcher can identify relevant documents, discard irrelevant ones, and conclude that a given set of documents sufficiently answers the question.

Adding to the complexity of model retrieval is the fact that the distributions of the variables and the correlations among the variables may differ from dataset to dataset, even if the datasets have similar metadata. This creates the need to specify the functional form of the relationships, and estimate the parameters of those functional forms for each data set. This process has no finite end. As Box and Draper (1987) pointed out, the modeler does not eventually arrive at the "correct" specification. Instead, the analyst can only achieve a "satisficing" model that balances the tradeoff between accuracy and complexity. The analyst knows it is time to stop refining the model when the ability to predict, manipulate or understand the data cannot be further improved in a cost effective manner. Therefore, the speed of this process depends on the skill of the modeler, the strength of the relationships among the data items, and the support that can be provided by a modeling environment.

Clearly model retrieval includes aspects of both data and document retrieval. But it also requires a level of manual intervention that is fundamentally different from either of these. Because model retrieval is such a complex process, any information system designed to facilitate model retrieval must be part of a larger, structured methodology for model formulation. This need for manual intervention, therefore, indicates that the model retrieval process cannot be considered complete until the intervention, i.e., the refinement of the retrieved model into a satisficing model, is complete.

Model Management Research

A great deal of work was done in the model management area during the 1980s and 90s. For example, Geoffrion (1987) identified two major problems confronting the management science/operations research (MS/OR) community. First, he noted that doing MS/OR tends to be a low productivity activity. Second, he noted that managers and policy makers are reluctant to ask for model-based assistance. Geoffrion, and others, tried to address these problems by developing modeling languages.

The modeling languages of the 1980s and 90s had four major design objectives. First, modeling languages should represent large and complex models using a few relatively simple statements (Geoffrion 1987; Brook et al. 1988; Fourer et al. 1990). Second, modeling languages should support the entire modeling life-cycle (Fourer et al. 1990; Geoffrion 1987, 1989). Third, modeling languages should allow the accumulation, sharing, integration, and reuse of data, models, solvers, and derived knowledge (Brooke et al. 1988; Choobineh 1991). Fourth, modeling languages should improve the productivity and managerial acceptance of MS/OR activities (Geoffrion 1987).

Several modeling languages were developed. These included structured modeling language (SML) (Geoffrion 1987), generalized algorithm for mathematical systems (GAMS) (Brooke et al. 1988), a mathematical programming language (AMPL) (Fourer et al. 1990), linear, interactive and general optimizer (LINGO) (Cunningham and Schrage 2004), structured query language for mathematical programming (SQLMP) (Choobineh 1991), and the subscript-free modeling language (SFL). The developers of SFL (Lin et al. 2000) state that "In SFL, the steps the decision maker must go through to formulate a model are the same steps that the decision maker must go through to understand the problem. This makes SFL very user friendly" (p. 615). However, neither SFL nor any of the other modeling languages was widely adopted by nontechnical managers, who continued to view MS/OR models as both confusing and expensive to build.

This is at odds with the notion of self-service BI. The *2014 State of Self-Service BI Report* (Logi Analytics 2014) notes that "Business users should be able to use all this information when they want, where they want, and do so without having IT in the way" (p. 3). Fifty-two percent of managers stated that it was important to have the capability to gain insight from data independent of their IT department (Logi Analytics 2014), but only 22% of the respondents actually have access to those tools now. The study also reports misalignment in priorities between IT and business departments. IT considers the use of spreadsheets to be the most important modeling for business users, whereas business users said it's most important for them to not only consume preformatted reports, but also to analyze data and create reports

on their own. Further, the report states that "the most important capabilities for business users were the ones they were the least satisfied with" (p. 3).

The model management work that was conducted in the 1980s and 90s failed to satisfy the desire of business managers for self-service BI tools. There are several reasons for this. First, the primary focus of prior model management research was how to build, store, and retrieve deterministic models. Second, the work assumed the modeler knew the relevant variables for the deterministic model, and was interested in finding the optimal solution to a structured problem. As pointed out by Davenport et al. (2001), business analytics deals with structured, semi-structured, and unstructured problems.

Modern modeling tools such as SAS Enterprise Miner and IBM SPSS Modeler allow for non-deterministic models and have graphical user interfaces that make them much easier to use, but these tools still are largely based on the SAS SEMMA process for modeling: sample, explore, modify, model, and assess (SAS Institute 1998; Rohanizadeh and Moghadam 2009). An implicit assumption of this approach is that the analyst knows the relevant variables to include in the model before they begin. This step is critical – Davenport (2013, p. 77) states "The essence of analytical communication is describing the problem and the story behind it, the model, the data employed, and the relationships among the variables in the analysis." Davenport and Kim (2013, p. 186) cite Intel Fellow Karl Kempf's statement that "effective quantitative decisions 'are not about the math; they're about the relationships.'" Effective self-service BI modeling tools must help managers determine which variables are relevant and the nature of the relationships between them.

Model management, at least in terms of reusing models, has been made easier through metadata management practices, languages, and standards. The sharing of data warehouses is greatly aided by the use of the Common Warehouse Metamodel (Object Management Group 2003). The ability to reuse BI models across different development platforms is enhanced by the use of the Predictive Model Markup Language (Guazzelli et al. 2009). These metadata approaches have made it easier for experts to share and reuse models, but will likely overwhelm the self-service BI users.

In addition, managers need help determining when to stop modifying a model and how to assess its usefulness. If managers do not receive help with assessing the usefulness of models, they may do more harm than good with models that they build. This raises the question of whether nontechnical managers ought to be building their own models. Pack (1987) recommends that an analyst have at least a master's degree in statistics, or the equivalent, in order to build and use forecasting models successfully. Geoffrion (1987) and Murphy et al. (1992) further argue that most modeling work is understood only by a small group of professionals, not nontechnical decision makers or managers. If that level of expertise is needed, then self-service BI will not be realized. However, Davenport (2013, p. 77) quotes Xiao-Li Meng, the chair of Harvard's Department of Statistics as saying:

> Intriguingly, the journey, guided by the philosophy that one can become a wine connoisseur without ever knowing how to make wine, apparently has led us to produce many more future winemakers than when we focused only on producing a vintage.

Apparently, as persons who did not know anything about wine making became involved in wine tasting, they also became more curious as to what creates the taste in wine. If managers are able to assess the usefulness of models, they also may become more interested in what allows their models to produce useful results. However, this will occur only if managers are confident that they can assess the usefulness of models. If self-service BI is to be realized, a methodology is needed that helps managers and business analysts throughout the SEMMA process; locate the relevant variables, see how those variables relate to each other, know when to stop modifying a model, and assess its usefulness.

The Model Formulation Process

As we've established, model formulation is a multi-step process based on the highly complex and open-ended task of model retrieval. The parameters generally are not known, and often the selection decision must be made without fully understanding the problem at either an individual or organizational level. In fact, the model formulation process can be characterized as a "wicked problem," as it has "unstable requirements and constraints based on ill-defined environmental contexts" (Hevner et al. 2004, p. 81). It also requires some degree of human intervention to arrive at a solution (Hevner et al. 2004). Therefore, model formulation can be looked at as a non-deterministic, problem-solving process – some models may be more useful than others, but there never is a definitive, "correct" model.

To support this formulation process, we propose a structured methodology that both technical and non-technical analysts can use to formulate analytical models (see Figure 1). While our approach does not reduce the complexity of the problem, it does provide a repeatable set of steps to approach model formulation. The steps are outlined below and demonstrated using the example of determining which incoming college students are least likely to graduate:

1) **Define the problem** by describing the decision to be made. In our example, the problem definition would be: "An inability to graduate on-time has added to the financial burden and accessibility of a college education. Which students are most likely to have difficulty graduating on time?"
2) **Determine the hypothesized relationships** that will inform the decision. This requires reducing the problem definition to a set of core concepts, such as retention, prior academic performance, and current working status.
3) **Define the data required** to test those relationships, specifically framed in terms of outcome (dependent) and input (independent) variables. In our example, an outcome variable is second-year retention and input variables include family income, high-school GPA, first-semester college GPA, and hours-per-week worked.
4) **Assess available data** to determine what data the decision-maker already has and what data they are capable of getting. Data quality should also be considered, as data might be available but useless for analysis. For example, family income and GPA data could be part of a student's existing record, but whether a student is working would likely require manual collection.
5) **Retrieve a set of candidate models** that would test the hypothesized relationships. All candidate models would have to be appropriate given the characteristics of the data (e.g., type, distribution). In our example, we might find that some have used regressions to build a predictive model of student success, while others have used clustering techniques to create profiles of high risk and low risk students. We may also find that several regression models have been used in the past with different subsets of the independent variables.
6) **Evaluate and refine the candidate models** arriving at the "best" final model for use in supporting the decision. We define "best" as an optimal tradeoff between accuracy and complexity, weighted according to the decision-maker's preferences. The decision-maker may test all candidate models and further refine them based on the characteristics of the specific data set. For example, for non-traditional student populations, high school GPA may be irrelevant.

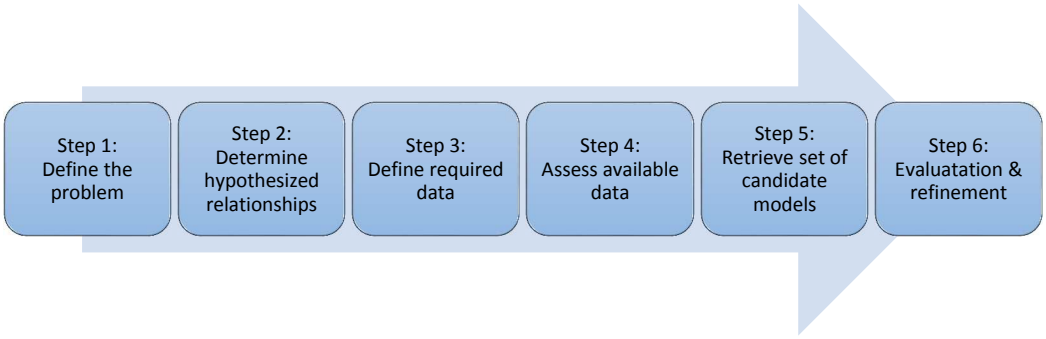

Figure 1. A Structured Methodology for Model Formulation

Dimensional Document Mart to Support Modeling: The Model Management Warehouse

While the methodology outlined in the previous section is useful in providing structure for the inherently open-ended model formulation process, it requires access to a sophisticated body of knowledge that encompasses data, statistical relationships, analytical modeling, and domain-specific organizational processes. Specifically:

- Business questions about organizational processes
- Data used to answer business questions -- variables to be explained/predicted (the *predictand* variables) and variables that influence the values of the predictand variables (the *predictor* variables)
- Hypothesized relationships among the predictor and predictand variables
- Mathematical representations of the hypothesized relationships among the predictor and predictand variables
- Measures of model effectiveness

We propose that a Model Management Warehouse, implemented as a dimensional document mart that stores each previously formulated model as a document, will facilitate model building in a way that is consistent with our model formulation methodology. In this document mart, the dimensions map to the aspects of an analytical model: the modeling domain, the predictand and predictor variables, variable transformations, the techniques to model the relationships among the variables, and measures of model effectiveness (see Table 2). Table 2 also describes each action a modeler takes during the process, including interactions with the dimensional store. Based on these dimensions we derive a star schema that can store the necessary data about the models (see Figure 2). In this schema, the fact table is "factless," and the result of any particular query is a set of documents that explains the models with the specified predictor and predictand variables.

Methodology Step	Implementation through Dimensional Data Store	Action Taken by Modeler
Step 1: Identify the business modeling domain	**Subject** dimension that identifies domains for business modeling – e.g., statistical profiling	Select relevant modeling domains for business problem being investigated
Step 2: Identify variables to be explained/predicted	**Predictand** dimension made up of keyword descriptors for the variable(s) that have been explained/predicted in prior models/studies – e.g., probability that an existing customer will drop your service	Select relevant predictand variables for current problem
Step 3: Identify variables that have been used in prior studies to explain/predict variable(s) of interest for this study	**Predictor** dimension made up of keyword descriptors for variables that were used in prior models/studies to explain/predict the predictand variables for this study – e.g., age, income, education	Select relevant predictor variables for current problem
Step 4: Identify possible analytic techniques for modeling the predictand	**Technique** dimension made up of keyword descriptors for broad analytic techniques that were used in prior studies to model the predictand – e.g., logistic regression, neural networks, decision trees	Select relevant techniques to use with data that is available for current problem

Step 5: Identify possible mathematical representations for hypothesized relationships between predictor and predictand variables	**Transformation** dimension made up of keyword descriptors for transformations that have been applied to the predictor and predictand variables in prior studies to improve the usefulness of the models – e.g., log, power, reciprocal, grouping, none	Select transformations that might improve the usefulness of models being considered
Step 6: Identify measures that can be used to assess the usefulness of models for the predictand variables	**Effectiveness** dimension made up of keyword descriptors for effectiveness measures for models being considered – e.g., lift, classification matrix, average precision	Determine values for effectiveness measures that indicate whether model is useful, and when it is time to stop the model building process.
Step 7: Retrieve model/studies that provide useful information for building your model	**Document** dimension made up of titles and associated URLs for documents that are relevant to your model building effort – e.g., documents that provide information for steps 1 through 6 above.	Abstract information needed to begin instantiation of models with your data

Table 2. Dimensional Document Mart to Support Model Formulation

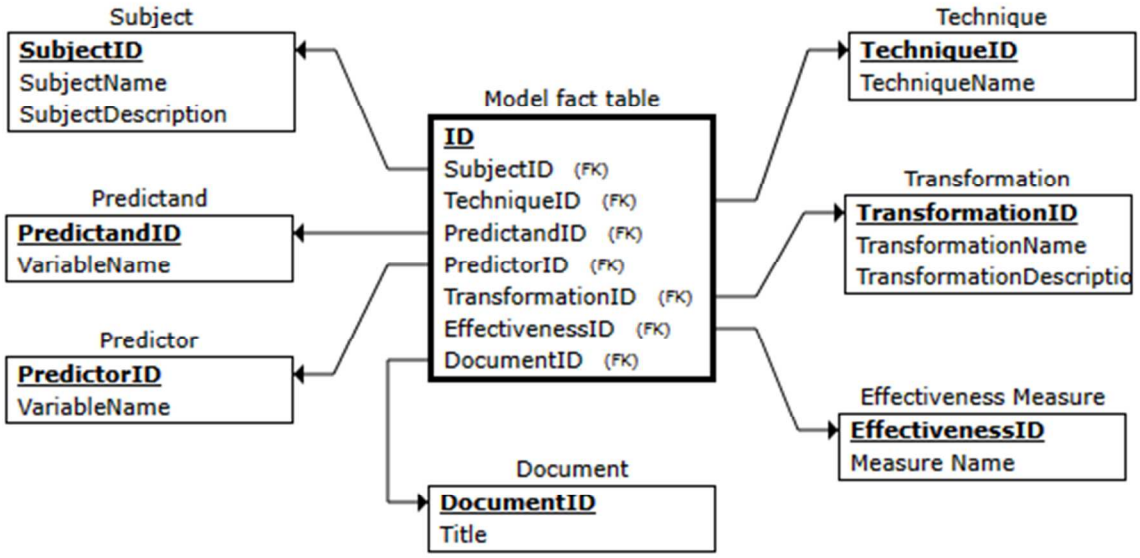

Figure 2. Star Schema for Dimensional Document Mart for Models

To illustrate how this document mart would be used to support the model building process, consider the following example:

> The state of Arizona wants to build a profiling model to identify Unemployment Insurance (UI) claimants that are likely to exhaust their benefits if they do not receive reemployment services. The business analyst tasked with building the model has a great deal of knowledge with respect to the UI program, but has only a modest amount of knowledge with respect to statistics.

A nontechnical business analyst would have a difficult time building this profiling model if the only tools available were SAS Enterprise Miner or SPSS Modeler. In order to begin the SEMMA process, the analyst must already have identified the relevant variables and collected data for those variables (SAS Institute 1998). In addition, the analyst must have some idea of how the variables relate to each other, what transformations might be useful, what statistical techniques might be helpful, and how to assess the usefulness of the models. It is unlikely that most UI business analysts would be able to do this on their own. In fact, a study conducted by the John J. Heldrich Center for Workforce Development (US Department of Labor) found that fewer than half of the UI jurisdictions had updated their profiling models since the great recession, and the primary reason for doing this was a lack of knowledge.

However, the business analyst *would* know that the modeling domain is "profiling" and the predictand variable is the "probability that a UI claimant will exhaust his or her benefits during the claimant's current spell of unemployment." With our proposed document mart, the analyst could select "Profiling Models" from the Subject dimension, and "Probability of Exhaustion" from the Predictand dimension. The US Department of Labor, several states, and various research organizations, such as Mathematica Policy Research and the Heldrich Center for Workforce Development, have published reports on the construction and use of profiling models to identify UI claimants that are likely to exhaust their benefits. If this information were placed in a dimensional document mart, constraining the dimensions using the problem-specific values of "profiling model" and "probability of exhaustion" would enable the analyst to see:

- All of the predictor variables (age, education, industry, income, etc.) that have been used to explain this predictand,
- All of the techniques (logistic regression, categorical models, neural nets, decision trees, etc.) that have been used to model this predictand,
- All of the transformations (logs, reciprocals, power functions, groupings, etc.) that have been performed on the predictor and predictand variables to improve the fit of the models,
- All of the measures (lift, percent classified correctly, classification matrices, etc.) that have been used to assess the usefulness of the models, and
- The degree of fit that has been acceptable to other modelers.

Moreover, if the analyst has data for only a limited set of variables, he/she also can filter on just that particular set of variables to see how well they have performed for other modelers. This information will enable business analysts to effectively use SAS Enterprise Miner or SPSS Modeler because it fills an essential gap between the problem definition, which is familiar to the analyst, and the modeling tool's graphical programming interface, which is becoming easier and easier for novices to use.

Discussion

As evidenced by the *2014 State of Self-Service BI Report* (Logi Analytics 2014), neither the model management work of the 1980s and 1990s, nor the current code-generating graphical interfaces of SAS and SPSS have enabled self-service BI. Model management research focused on developing modeling languages designed to enable operations research and management science (OR/MS) researchers – i.e., "experts" – to quickly build, store, retrieve, and reuse models. The two major shortcomings of this approach are: 1) it requires modelers to have a deep understanding of statistical modeling; and 2) it assumes the exact same instantiation of a model will be used multiple times. Given these shortcomings, it is not surprising that the systems did not achieve widespread use.

The code-generating interfaces of SAS Enterprise Miner and SPSS Modeler are promising – they greatly reduce the time required to learn the software and speed up the model-building process. However, like previous attempts at simplified modeling languages, they presume a level of mathematical, statistical, and modeling knowledge that is not present in most business analysts. As illustrated in our UI exhaustion example, a typical business analyst may not be able to identify the relevant predictor variables, and probably is even less certain about how to transform variables, select a technique for modeling the relationships among the variables, or assess the usefulness of the model. SAS's SEMMA process provides no assistance with selecting the predictand and predictor variables, and provides very little guidance with respect to transformations, technique selection, or assessment.

Kimball (1997) argues that dimensional modeling is the only viable technique to support end-user queries in a data warehouse. We argue that dimensional document marts are the only viable technique to support self-service BI. The intuitive structure of a dimensional data mart enables users who know little about databases or query languages to get the information they need. The widespread availability of tools such as Excel provide a low-cost way of connecting users with these dimensional databases. This has contributed greatly to the widespread use of dimensional data marts.

The same is true for dimensional document marts to support model-building. Once the fact and dimension tables are constructed, Excel's pivot table functionality can query the database. The user merely has to click on a specific dimension (or "slicer"). The process of constructing the dimension and fact tables from a collection of text-searchable documents can also be automated through software, detecting keywords from a list of modeling domains, predictands, predictors, techniques, transformations and effectiveness measures. This makes it possible to maintain and update the dimensional document warehouse with little or no human intervention.

In the case of dimensional data marts, nontechnical business analysts clearly are able to understand the data and performance measures. Their domain-specific knowledge about the organization makes them more qualified than technical staff to interact with the data. The obstacle for nontechnical analysts, prior to dimensional data marts, was simply access to the data. Figure 3 shows a possible mockup of the output from a query, depicted as a Pivot Table in Microsoft Excel, to help the analyst select the predictand variables and the predictor variables for the states of Arizona, Arkansas, and California.

Independent Variables	Dependent Variables
Select	Count of Binary_Exhaust
Delay_in_Filing	1
Education	3
Industry	2
Job_Tenure	2
Maximum_Benefit_Amount	1
Occupation	1
Potential_Duration	1
Residence	1
Wage_Loss_Replacement	2

State_Name

Arizona

Arkansas

California

Figure 3. Result Set from Dimensional Model Management Warehouse Query

The information in Figure 3 clearly will be useful to the modeler. The pivot table lists the nine predictor variables used in models that attempt to predict the probability of exhaustion in three states. It also shows how many models used each predictor variable. Further queries would show the statistical technique used by each model (logistic regression for Arizona, multiple linear regression for Arkansas, and a neural network for California), the transformations that were applied to each variable by each model to improve fit (e.g., groupings, power function, logs), the measures used to assess the fit of each model (classification matrices and lift charts), and how accurate the models were when each model's building process was stopped.

This paper outlines a methodology and technology artifact that provides modelers with the key pieces of information required to execute the SEMMA process. It is clear that providing this information is a necessary condition for self-service BI, but whether it is enough to give nontechnical managers the ability

to independently construct useful models must be tested. The next step is to test this proposition through a field experiment using a full prototype. Future research will also explore the viability of automatic code generation based on analysts' selection of predictors and predictands, variable transformations, statistical procedures, assessment measures, and the stopping rule. This would be the true "missing link" in self-service BI, all but eliminating the need for the nontechnical analyst to be even moderately proficient with statistical packages such as SAS and SPSS.

REFERENCES

Blair, D. C. 2002. "The Data-Document Distinction Revisited," *The DATA BASE for Advances in Information Systems* (37:1), pp. 77-96.

Box, G. E. P., and Draper, N. R. 1987. *Empirical Model Building and Response Surfaces*, New York, Wiley.

Brooke, A., Kendrick, D., and Meeraus, A. 1988. *GAMS: A User's Guide*, Redwood City, CA: Scientific Press.

Choobineh, J. 1991. "SQLMP: A Data Sublanguage for Representation and Formulation of Linear Mathematical Models," *ORSA Journal on Computing* (3:4), pp. 358-375.

Cunningham, K., and Schrage, L. 2004. "The LINGO Algebraic Modeling Language," in *Modeling Languages in Mathematical Optimization*, J. Kallrath, (ed.), Kluwer Academic Publishers, pp. 159-171

Davenport, T. H. 2013. "Telling a Story with Data," *Deloitte Review* (12).

Davenport, T. H., Harris, J. G., De Long, D. W., and Jacobson, A. L. 2001. "Data to Knowledge to Results: Building an Analytic Capability," *California Management Review* (43:2), pp. 116-138

Davenport, T. H., and Kim, J. 2013. *Keeping UP with the Quants*, Harvard Business Review Press.

Fourer, R., Gay, D. M., and Kernighan, B. W. 1990. "A Modeling Language for Mathematical Programming," *Management Science* (36:5), pp. 519-554.

Geoffrion, A. M. 1987. "An Introduction to Structured Modeling," *Management Science* (33:5), pp. 547-588.

Geoffrion, A. M. 1989. "The Formal Aspects of Structured Modeling," *Operations Research* (37:1), pp. 30-51.

Guazzelli, A., Zeller, M., Lin, W-C., Williams, G. 2009. "PMML: An Open Standard for Sharing Models," *The R Journal* (1:1). p. 60.

HBR Analytic Services. 2012. "The Evolution of Decision Making: How Leading Organizations Are Adopting a Data-Driven Culture," Harvard Business Review (online). https://hbr.org/resources/pdfs/tools/17568_HBR_SAS%20Report_webview.pdf retrieved 2-14-2015.

Henschen, D. 2014. "IBM Watson Analytics Goes Public," InformationWeek. http://www.informationweek.com/big-data/big-data-analytics/ibm-watson-analytics-goes-public/d/d-id/1317887 retrieved 2-14-2015.

Hevner, A. R., March, S. T., Park, J., Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 77-105.

Kimball, R. 1997. "A Dimensional Modeling Manifesto," *Database Magazine* (10:9), pp. 59-78.

Kottemann, J. E., and Dolk, D. R. 1992. "Model integration and modeling languages," *Information Systems Research* (3:1), pp. 1-16.

Lin, E., Schuff, D., and St. Louis, R. 2000. "Subscript-Free Modeling Languages: A Tool for Facilitating the Formulation and Use of Models." *European Journal of Operational Research* (123:3), pp. 614-627.

Logi Analytics. *2014 State of Self-Service BI Report*. http://images.learn.logixml.com/Web/LogiAnalyticsInc/%7B7c21cd62-221c-44af-9ecd-a35265bc8e34%7D_LogiAnalytics-2014StateOfSelfService-Artwork-1028.pdf retrieved 2-8-2015.

Murphy, F. H., Stohr, E. A., and Asthana, A. 1992. "Representation Schemes for Linear Programming Models," *Management Science* (38:7), pp. 964-991.

Object Management Group. 2003. Common Warehouse Metamodel (CWM) Specification. http://www.omg.org/spec/CWM/1.1/PDF/ retrieved 4-24-2015.

Pack, D. J. 1987. "A Practical Overview of ARIMA models for Time Series Forecasting" in *The Handbook of Forecasting: A Managers Guide*, S. G. Makridakis and S. C. Wheelwright, (eds.), New York: Wiley, pp. 196-218.

Rohanizadeh, S., and Moghadam, M. 2009, "A Proposed Data Mining Methodology and its Application to Industrial Procedures" *Journal of Industrial Engineering* (4), pp. 37-50.

SAS Institute. 1998. "Data Mining and the Case for Sampling," SAS Institute Best Practices Paper, Carey, NC.

US Department of Labor. Unpublished study conducted by John J. Heldrich Center for Workforce Development for the US Department of Labor.