

## **REVELATIONS FROM DATA MINING: A CASE STUDY OF A SALES TERRITORY**

*By Alan J. Dubinsky, Guoying Zhang and Michele Wood*

The economic climate and extensive product proliferation have created intense competition for most sales personnel. Salespeople must out-maneuver competitors and strategically target appropriate products and services to fit their customers' needs. In such a difficult business milieu, data mining for business intelligence can provide salespersons a treasure trove of information with which to attend to their customers' unique requirements.

This paper presents a data mining process that salespeople can utilize to analyze customer records to help them better understand customers' purchasing tendencies and associations. The case study can be utilized to tailor a salesperson's efforts by targeting specific customers based on their characteristics. Customer sales data from a territory of a large industrial distributor were employed to apply three different data mining techniques. Data mining results yielded valuable information that could be used to increase territory sales, gain customer loyalty, and grow market share.

### **INTRODUCTION**

Customers are the lifeblood of any firm (Dubinsky 1999): they nourish the company financially. Indeed, Kale (2004) posits that developing and maintaining customer relationships is vital for competitive advantage. Failure to pay heed to customers and their various needs can leave a marketer unable to leverage the company's competencies. In an effort to provide customers with the "perfect customer experience," Customer Relationship Management (CRM) has become requisite for many organizations as a means of enhancing performance (Javalgi, Martin, and Young 2006; Payne and Frow 2005, 2006).

Successful CRM implementation requires attention to four crucial areas: strategy, people, processes, and technology (Crosby and Johnson 2001; Yim 2002). Operationalizing this perspective, Yim, Anderson, and Swaminathan (2004) propose that implementation entails focusing on key customers, organizing around

CRM, managing knowledge, and incorporating CRM-based technology. CRM is an "enabler," as it assists companies to manage their customers effectively (Landry, Arnold, and Arndt 2005; Moutot and Bascoul 2008; Raman, Wittman, and Rauseo 2006).

Boulding et al. (2005) state that the core of CRM is "dual creation of value." Through CRM activities, value is created for both the marketer and the customer. That is, rather than solely trying to maximize value for the firm, management should also be concerned with creating value for the customer. Value creation entails discerning what value the organization might offer customers, ascertaining what value customers provide the firm, and maximizing the lifetime value of the customer (Payne and Frow 2005).

Customer lifetime value (CLV) is the predicted profitability of a customer during the entire relationship with the company (Kale 2004). CLV changes over the life of the relationship and

requires integration of the organization with the entire CRM process (Boulding et al. 2005). CRM efforts necessitate an ongoing balance between the customer and the firm as the relationship evolves through its stages (Reinartz, Krafft, and Hoyer 2004). Moreover, distribution of value across customers is uneven (Chen et al. 2006; Reinart, Krafft, and Hoyer 2004). Therefore, marketers must balance their customer portfolios vis-à-vis their customers' needs, company efforts to foster customer value, and the means by which customers provide value for the firm (Boulding et al. 2005). Part of this entails correctly identifying responsive and profitable target customers (e.g., Cao and Gruca 2005).

A key goal of CRM is to efficiently and effectively increase the acquisition and retention of customers by selectively building and maintaining mutually satisfying relationships with them. As Dickson et al. (2009, p. 113) aver: "When CRM is done right, it seamlessly integrates the marketing-sales-delivery-customer-service core functional added value processes...." Through advances in information technology, companies can realize improved customer relationships. CRM technology can foster gathering, analyzing, and interpreting various kinds of customer data in order to enhance the relationship with the customer. By converting customer information into usable data, CRM can increase the overall performance of a company (Yim, Anderson, and Swaminathan 2004; Stein and Smith 2009).

Use of information technology to manage customer relationships is a critical dimension of relationship marketing (Anderson, Dubinsky, and Mehta 2007). In fact, it has been found to enhance salesperson performance (Hunter and Perreault 2007; Mathieu, Ahearne, and Taylor 2007; Stoddard, Clopton, and Avila 2006) and

other CRM outcomes (Moutot and Bascoul 2008). Although diffusing technological innovation into a sales force may face resistance (Sharma and Sheth 2010), providing technology empowerment assists in salespeople's technological usage (Chou, Pullins, and Senecal 2007).

A particularly beneficial technological tool for transforming customer information into usable data—and thereby assisting sales personnel to build sound relationships with customers—is data mining. Data mining refers to "all aspects of an automated or semi-automated process for extracting previously unknown or potentially useful knowledge and patterns from large databases" (Olafsson, Li, and Wu 2008, p. 1429). Its main objective is to reveal interesting patterns, associations, rules, changes, irregularities, and general regularities from the data to improve decision making (Sumathi and Sivanandam 2006, p. 321).

Data mining can help sellers (including sales personnel) enhance their understanding of customers (Shaw et al. 2001, p. 128), as it entails "...searching and analyzing data in order to find implicit, but potentially useful, information...to uncover previously unknown patterns, and ultimately comprehensible information from large databases." Furthermore, data mining provides specific knowledge to organizations (e.g., salespeople) about each customer's needs (Shaw et al. 2001). Moreover, it allows users to identify relationships between independent and dependent variables (Zhang et al. 2008). In addition, data mining approaches can reduce the effect of human bias on data analysis (Tu 2008).

Data mining can reveal a treasure trove of customer information and assist in a firm's knowledge management process (Shaw et al. 2001). This applies whether data mining is

conducted at an aggregate company level or on an individual sales territory. Consequently, it can serve as an especially effective tool to help a salesperson attend to the unique needs of each of his or her customers. Moreover, it affords enhanced capability to deal with each customer. As Shaw et al. (2001, p. 128) assert:

There is an increasing realization that effective customer relationship management can be done only based on a true understanding of the needs and preferences of customers...[d]ata mining tools can help uncover the hidden knowledge and understand customers better.

Given the important role data mining can play in enhancing relationship marketing efforts, salespeople should be knowledgeable about how to perform data mining. They must possess the technological tools and know-how to conduct data mining in a way that will allow them to leverage such efforts to advantage. Indeed, Anderson, Dubinsky, and Mehta (2007, p. 121) state that data mining is an especially useful tool for sales personnel, as it can "identify important relationships among buyers and the products they purchase, along with associated complementary products" and help salespeople assist their customers to sell effectively through their own accounts.

The current paper presents a non-convex, useful approach to data mining that has actually been utilized by a salesperson in an industrial setting. Discussed are three data mining techniques, each with a unique focus that the host salesperson conducted with data from her accounts. The case study that follows is an attempt to show salespeople the value that their own use of data mining can offer. As such, this paper partially answers a call by Raman, Wittmann, and Rausea (2006, p. 51) for "...future research [that] could focus on fine-grained aspects of CRM implementation and use."

## **DATA MINING ANALYSIS**

In general, the data mining process includes the

following stages: data preparation and exploration, data training and validation, model evaluation and selection, and application of the model on new data (Berson et al. 1999). In the stage of data training and validation, a number of alternative methods exist. They can be categorized into the following four groups: clustering, classification, regression, and association rule learning (Fayyad, Piatetsky-Shapiro, and Smyth 1996). Shown in Table 1 are these four data mining techniques with their corresponding purposes.

The objective of this paper is to illustrate a data mining process that salespersons could use to develop a grow-sales strategy. This is done through examining relationships among various customer variables and sales information that are conceivably already in a firm's data base. Using Table 1 as a guide, three data mining methods were employed in the present paper: multiple linear regression, association rules, and clustering. Regression analysis was used to develop models to predict annual sales revenue based on customer characteristics. Association rules were generated for the purpose of finding products that customers often purchase together; such information would be helpful for developing cross-selling strategies. Clustering was conducted to profile customers and develop corresponding marketing strategies for different clusters (i.e., related groups of customers)<sup>1</sup>.

The remainder of this paper will illustrate how an industrial salesperson undertook the three foregoing data mining techniques. Specific details of how the host salesperson conducted data mining will be described, thus providing user-friendly directions for sales personnel interested in utilizing data mining within their own territories.

Table 1. Common Data Mining Methods	
Technique	Purpose
Clustering	Discovering groups and structures in the data that are in some way “similar.”
Classification	Making classifications of new data with known classes. Common algorithms include decision tree, nearest neighbor, naive Bayesian, neural networks, and support vector machines.
Regression	Estimating parameters in an assumed function form by fitting the data with the least error. The established function can be used to predict new data.
Association Rule	Searching for association relationships between variables. A common example is referred to as “market basket analysis” to determine products or services that will be purchased together.

### Data Collection

Initially, salespeople must identify and collect the data that they feel is pertinent to their specific data mining situation. Sales personnel interested in data mining ideally should be able to obtain from company records the data utilized in this paper. Indeed, the host salesperson in this study requested the data from her management in order to undertake data mining. She focused solely on her own accounts and described the importance of data mining to management to justify her being given the data. When proposing to management the advantages of using data mining, salespeople can make the following arguments:

- Data mining can reveal an abundance of customer information.
- Data mining helps salespeople improve their understanding of customers.
- It requires users to search and analyze data to find useful information that can uncover previously unknown patterns of relationships among customer variables (e.g., sales volume and such factors as customer size, channel of distribution, use of e-commerce).
- Data mining can offer detailed information about each customer’s needs.
- It allows firms to make enhanced use of the information they have about their customers.

- Data mining can serve as an especially effective tool to help salespeople attend to the unique needs of their customers.

As such, the salesperson must first identify the relevant group of accounts on which he or she wishes to undertake data mining. The salesperson in our case study gathered data for a 12-month period from an average-sized sales territory of Company DM (the real name of the company is not disclosed for confidentiality). DM is a *Fortune 500* company and one of North America’s largest maintenance, repair, and operations (MRO) product distributors. It competes with other broad-line distributors, regional suppliers, and local suppliers. The host salesperson’s entire sales territory consists of twenty customers in the commercial manufacturing and health care segments<sup>2</sup>. The average annual sales volume for the host salesperson’s customer data set was \$103,487.56. The standard deviation was \$23,497.83. The smallest customer generated annual sales of \$4,477.86 (a new health care customer), and the largest customer in the data set purchased products totaling \$330,513.00 over the 12-month period.

### Data Preparation

After the salesperson has identified the relevant

**Table 2: Product Category Spending by Selected Customers**

Customer	LAMPS	BALLASTS	FIX-TURES	PORTABLE LIGHTING	FLASH-LIGHTS/	ELECTRICAL SUPPLIES
1	\$86					\$82
2	\$596		\$6	\$364	\$2,700	\$706
3		\$29	\$344			\$1,298
4	\$17,156	\$61	\$12,281	\$489	\$1,076	\$44,541
5				\$141		
6			\$5,190		\$25	\$1,904
7	\$5,252	\$1,205	\$13,611	\$3,298	\$3,416	\$40,797
8	\$2,181	\$276		\$136		\$2,275

accounts for data mining, he or she should classify the data that will be used in the analyses. The classification could be at two levels—product grouping and customer profiles.

**Product Groupings** In a multi-product firm, the salesperson would need to determine how the product mix can be meaningfully classified (e.g., product line, complementary items, individual items). With respect to product grouping in our case study, the host salesperson examined her product sales for each customer, ultimately categorizing the sales into the 102 distinct DM product categories. Shown in Table 2 is the annual spending on selected product categories for a *portion* of the host salesperson's twenty customers during the 12-month period.<sup>3</sup>

**Customer Profiles** In addition to grouping products, the salesperson needs to identify germane characteristics that can be used to classify customers vis-à-vis the nature of their interactions with the firm. Doing so will produce a profile for each customer. Prospective characteristics might include such features as customer size, customer

demographics, purchase patterns, account call frequency, and so on. For each characteristic, the salesperson will assign a numerical value (e.g., 0, 1, 2) to represent the corresponding category of the characteristic.

With respect to customer profiles in our case study, the host salesperson identified certain customer dimensions of her twenty customers that could meaningfully portray her accounts. Based on her knowledge of the territory and input from DM management, the following profile characteristics of each customer, as portrayed in Table 3, were selected:

1. Customer Segment Type: 0=Manufacturing, 1=Health care.
2. Corporate Account: 0=No, 1=Yes, (i.e., whether the customer has corporate account status with DM's corporate office).
3. VMI: 0=No, 1=Yes. VMI refers to "vendor managed inventory." This means that a DM employee goes to the customer location weekly with a bar code scanner to order products, place replenishment product

onto customer shelves, and assist the customer with product selection for spot (unplanned/impulse) purchases.

4. PPD (*Pre-paid*) Freight: 0=No, 1=Yes. DM pays the freight for customers whose purchase level reaches a particular threshold; some of DM's corporate contracts receive this incentive.
5. E-commerce Customer: 0=No, 1=Yes. This variable indicates whether the customer utilizes some kind of electronic means (e.g., e-mail, EDI) to conduct transactions with DM.
6. Distance from DM Store: Total distance in miles from the closest DM store.
7. Primary Order Method: Three different primary order methods were used, coded as follows: 00 = Phone, 01 = Fax, 10 = E-commerce.
8. MRO Potential: Total annual dollars that the customer spends with all suppliers for maintenance, repair, and operating products.
9. Number of Contacts: Total number of individuals within the customer's firm who order from DM.
10. E-commerce Share: The percentage of the customer's sales that are procured through electronic channels (i.e., EDI, DM.com, email, or some other type of electronic platform).
11. Primary Discount: The minimum discount that the customer receives for any item purchased from the DM catalog.
12. Count of Categories Purchased: The total number of distinct product categories from which the customer purchased items.

## DATA ANALYSES

At this juncture in the process, the salesperson now selects alternative data analytic techniques that he or she could use for data mining. The techniques ultimately used should be based on the purpose for undertaking the data mining. As noted earlier, shown in Table 1 are the four general data mining techniques and their corresponding purposes. In our cases study, multiple linear regression, cluster analysis, and association rules were the techniques selected. The host salesperson opted for these three techniques because her purposes for data mining were to (a) employ a set of customer characteristics to predict sales revenue per customer (thus, multiple linear regression), (b) segment the territory into customer groups (thus, cluster analysis), and (c) identify cross-selling opportunities (thus, association rules). Each of these techniques will now be described vis-à-vis salesperson data mining.

### Multiple Linear Regression.

If a salesperson wishes to use multiple linear regression in data mining, most likely he or she will utilize it to forecast sales revenue per customer. Multiple linear regression is a prediction-type analysis where historical data are used to build a model that helps forecast a future event (i.e., sales in this study). Accordingly, the salesperson must identify a set of germane customer characteristics that are likely to be related to sales revenue. The customer profile that the salesperson has already created (as noted above) should conceivably consist of those relevant customer characteristics.

Once the requisite data have been obtained for use in multiple linear regression, the salesperson

Table 3. Customer Characteristic Data

Customer	12-month Sales	Customer Type	Corporate Account	VM I	PPD Freight	E-Commerce	Distance	Primary Order Method (e-commerce)	Primary Order Method (Fax)	MRO Potential	Number of Contacts	E-commerce Share	Primary Discount
1	\$25,000.00	0	0	0	0	1	20	1	0	2000000	27	45.57	5
2	\$108,091.26	0	0	0	1	1	50	1	0	2000000	42	57.62	10
3	\$16,464.12	1	1	0	0	0	150	0	0	200000	14	1.15	10
4	\$330,513.00	0	1	1	1	0	65	0	1	800000	31	0	15
5	\$37,967.80	0	0	0	0	0	100	0	0	900000	11	0	5
6	\$43749.00	0	0	0	0	0	12	0	1	1300000	30	17.12	10
7	\$220,556.48	0	0	0	1	1	15	1	0	200000	82	35.86	10
8	\$34,781.97	1	1	0	1	1	75	1	0	300000	8	89.7	10
9	\$19,962.18	0	0	0	0	0	75	0	0	2000000	17	1.95	7.5
10	\$289,831.27	0	1	0	1	1	45	1	0	300000	32	10.23	10
11	\$46,821.18	1	1	0	1	1	7	1	0	80000	5	94.35	10
12	\$70,000.00	0	0	0	0	0	12	0	0	400000	13	0	10
13	\$51,417.57	0	0	0	0	0	20	0	0	700000	12	0	10
14	\$65,934.76	0	0	0	0	0	5	0	0	100000	8	1.11	10
15	\$4,477.86	1	1	0	0	1	75	0	0	40000	8	0	10
16	\$213,149.00	0	1	1	1	0	80	0	0	1400000	9	0.89	13
17	\$13,678.01	0	0	0	0	1	50	1	0	600000	17	43.21	5
18	\$192,790.40	0	1	1	1	1	1	0	0	700000	43	0	10
19	\$258,902.14	0	0	0	1	1	5	1	0	1100000	31	47.6	15
20	\$25,663.24	0	0	0	0	1	5	0	1	700000	15	30.09	5

**Note:** Phone orders were coded as “00”. Therefore, if both columns for “Primary Order Method” (e-commerce and fax columns) contain a “0” for a particular customer, this means that the customer was using the *phone* as the *primary* order method.

must select computer software to conduct the analysis. An array of software options is available on the market. Prototypical options include SAS (Statistical Analysis Software), SPSS (Statistical Programming for Social Science), STATA, and XLMiner for Excel. Each software package provides essentially similar routines for multiple linear regression.

A primary research question of our host salesperson was, "What factors (those noted in Table 3) determine the sales volume of a customer?" For example, does giving the customer pre-paid freight increase sales? Does deploying labor in the form of vendor managed inventory augment sales? If the customer's corporate office enforces the contract, might sales rise? The host salesperson's goal was to understand what DM programs and services, if any, may be used to forecast a customer's annual sales. Therefore, the data set presented in Table 3 was utilized for regression analysis. The resulting model could then be applied to develop and implement a sales plan directed at enhancing sales with other customers in the territory.

Results from the regression analysis are presented in Table 4. The host salesperson utilized Excel to conduct the analysis owing to its widespread availability in Microsoft Office. The multiple *R*-square of 0.99 indicates that the regression model explains 99 percent of the variation in total sales dollars of the customers in the study. This means that the independent variables (customer profile characteristics—Table 3) chosen for this regression analysis are helpful in explaining the variance in the dependent variable—annual sales. The significant independent variables—as evidenced by the statistically low *p*-value (less than 0.05)—are corporate account, pre-paid freight, primary order method/e-commerce, primary order

method/fax, and E-commerce share.<sup>4</sup>

The regression model can be interpreted as follows: First, if a customer is a corporate account, he or she might generate \$65,681.04 *less* sales revenue (owing to the negative sign of the value: -65681.04) than a customer who is not a corporate account. This implies that DM's sales program which offers *local* target customers deeper discounts than corporate customers seemingly yields unexpected results. The program apparently enhances sales from local customers but not from corporate accounts. Second, if a customer is accorded prepaid freight, annual sales from that customer could well be \$182,378.35 higher than a customer who does not receive prepaid freight. Therefore, prepaid freight may be an important variable for augmenting account sales. Third, a customer who utilizes an e-commerce channel for ordering (such as DM.com, EDI, or some other type of electronic interface) tends to have sales revenue that is \$121,670.15 greater than a customer who orders primarily via the phone. Similarly, a customer who faxes orders to DM seems to have \$51,957.55 more sales revenue than a customer who primarily phones in the orders. Finally, every unit increase in a customer's *E-commerce share* could well lead to a sales *decrease* of \$3,020.12 (owing to the negative sign of the value: -3,020.12). Given the popularity of using the internet on which to make orders, this result is not intuitive. Discussions with the host salesperson, as well as in-depth analysis of each account's sales data, however, revealed a potential rationale for the unexpected finding. By and large, the territory's accounts are not using e-commerce extensively with which to place an order. However, there exist customers who generate a relatively small level of sales while having a disproportionately high percentage of e-commerce. Accordingly, the



Table 4. Multiple Linear Regression	
<i>Regression Statistics</i>	
Multiple R	0.995
R Square	0.990
Adjusted R Square	0.967
F-statistic	43.717***
Observations	20
*** p<0.001	

	<i>Coefficients</i>	<i>Standard Error</i>
Intercept	15964.84	40009.62
Customer Segment Type	-2901.01	70321.93
Corporate Account	-65681.04*	26186.06
VMI	14911.31	25466.23
PPD Freight	182378.35**	45172.09
E-Commerce Customer	16325.02	22359.10
Distance from DM Store	70.58	308.19
Primary Order Method (e-commerce)	121670.15*	35535.82
Primary Order Method (fax)	51957.55*	17764.02
MRO POTENTIAL	-0.02	0.02
Number of Contacts	-72.45	390.30
E-commerce Share	-3020.12*	864.93
Primary Discount	4849.27	5059.43
Count of Product Categories Purchased	216.60	838.95
**p <0.01, *p <0.05		

host salesperson confirmed that she needs to expend more effort to encourage e-commerce in her territory.

### Cluster Analysis

A prominent use for data mining is to create underlying customer segments within a salesperson's territory. More specifically, it entails reducing the number of customers in a territory into a manageable set of clusters so that corresponding sales plans can be directed at each

cluster. This reduction occurs by using the customer profiles (identified above). If a salesperson's customers are segmented based on customer profile characteristics, then selling efforts could be tailored to each segment.

A variety of clustering techniques, as well as software providing cluster analyses, is available.<sup>5</sup> Regardless of the technique or software used, however, cluster analysis is a data reduction technique seeking similarity within a customer segment but dissimilarity across customer

segments. In other words, customers within a segment should have similar profile characteristics but have dissimilar profile characteristics from customers in other segments.

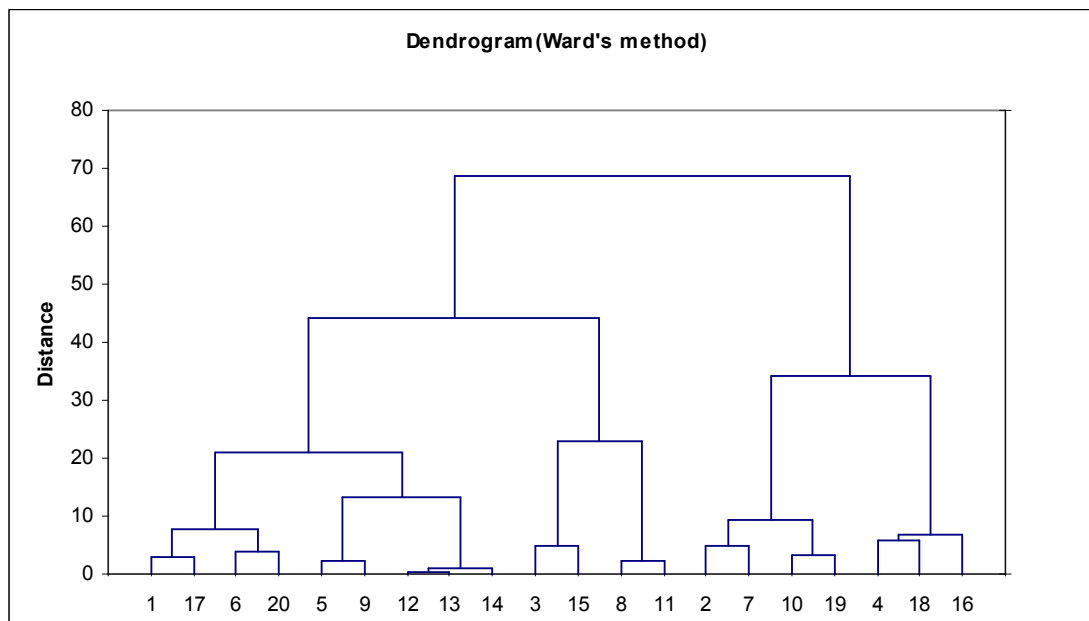
With respect to our case study, data for cluster analysis were from the same data set used for multiple linear regression. Shown in Figure 1 are the results from the host salesperson's cluster analysis. The clustering appeared to be accurate vis-à-vis identifying distinguishing characteristics that grouped the host salesperson's territory customers together. For example, customers 1 and 17 were manufacturers that had the following similarities (noted in Table 3): non-corporate account, non-VMI program account, no pre-paid freight, primary order method of e-commerce, e-commerce share of sales, and level of primary discount. Also, customers 6 and 20 were manufacturers having these overlapping characteristics: non-corporate account, non-VMI program account, no pre-paid freight, and primary order method/fax. A higher level of

clustering further grouped these four accounts together.

Another example is the cluster containing customers 4, 18, and 16. All three of these customers were manufacturers who generated relatively high sales volume, participated in the VMI program, received pre-paid freight, rarely used e-commerce for order-placement, had large sales potential, and purchased many different categories of DM products. A higher level of clustering further grouped these three accounts together with customers 2, 7, 10, and 19, all of whom possessed similarities in certain characteristics (noted in Table 3).

One further example entails customers 3, 8, 15, and 11, who ultimately form a cluster at a higher level. What did accounts 3 and 15 have in common? They both were health care institutions generating a relatively small sales volume, were corporate accounts not participating in the VMI program, were not accorded pre-paid freight, were a relatively far distance from a DM store, rarely utilized e-

**Figure 1. Dendrogram (Ward's Method) of Customer Segmentation**



commerce to place orders, had a small number of buyers of DM products within their organizations, and received the same primary discount. Commonalities between customers 8 and 11 were as follows: health care facility, corporate account with no participation in the VMI program, pre-paid freight recipient, e-commerce ordering, small number of buyers of DM products within the institution, and identical primary discount. Results from the cluster analysis (Figure 1) also accurately depicted similarities across other clusters.

### **Association Rules**

Another frequently used data mining technique is association rules (Shaw et al. 2001). This method is also known as affinity grouping or market basket analysis. When a salesperson uses affinity grouping, it is to determine which products are purchased together. A common example for this method is to determine which retail items go together in a shopping cart at the supermarket. An association rule has a generic format of the following: If product A is purchased, then product B should also be purchased. Product A is often referred as the “antecedent product,” and product B is the “consequent product.”

In order to derive a legitimate association rule, three critical parameters will be generated from the association rules software—namely, support, confidence, and lift. Support is the number of transactions for a particular customer that included both the antecedent and consequent products in the rule. Confidence is defined as the probability of purchasing a consequent product given that the antecedent product was purchased. Lift is defined as the ratio of confidence to the probability of purchasing the consequent product. According to Shmueli et al. (2007), a lift ratio greater than 1.0 suggests that an association is significant between antecedent

and consequent products. In other words, the chance of purchasing the consequent product given that the antecedent product was purchased is higher than the general chance of purchasing the consequent product. The larger the lift ratio, the greater is the strength of the association (of the antecedent and consequent products).<sup>6</sup>

With respect to our case study, the host salesperson’s third purpose of data mining was to identify the potential for the host salesperson to engage in cross-selling—selling an additional product or service to an existing customer. Association rules generated from various DM product categories could reveal which products tended to be purchased together. Armed with this knowledge, the host salesperson may be able to enhance sales in her territory via cross-selling (Kumar 2002).

Because not all 102 product categories were purchased by the customers in the host salesperson’s territory during the 12-month period, data for the association rule analysis falls into a subset of the total 102 DM product categories. The objective of the analysis was to find common associations between product categories that are both purchased together by a significant number of customers.

Prior to conducting association rules, data transformation had to be conducted for use in the association rules analyses. DM is a broad line distributor with a multiplicity of products. Based on customer sales records, a binary value (e.g., 0 or 1) was assigned for each customer in each product. For brevity, illustrated in Table 5 is a subset of the transformed data. When a cell value equals “1,” it indicates that the customer in that row purchased the product in that column in the past 12 months. If the cell value equals “0,” it means that no such transactions could be found in the data set for that customer for the

particular product.<sup>7</sup>

The host salesperson's transformed data were analyzed using association rules. In order to generate a reasonable size of legitimate rules, the minimum support and lowest confidence level had to be determined. After testing alternates, a minimum support of 4 and the lowest confidence level of 50% were chosen because these rules were generated with adequate lift ratios. Shown in Table 6 are the association rules generated when using a minimum support of 4—which means there needs to be at least 4 customers having the purchasing combination of the particular antecedent and consequent products. The analysis also set the lowest confidence level at 50%, which means that the probability has to be at least greater than 0.5. The findings revealed interesting product combinations that the host salesperson did not necessarily anticipate. Such revelations indicate the value association rules can offer data mining users.

According to the rules generated (Table 6), customers who bought lamps from DM also typically purchased safety products. (The

confidence level of this rule was 80%, and the lift ratio was 1.33.) Rule 2 shows that customers who purchased tools and storage products also bought safety products (with a 75% confidence level and 1.25 lift ratio). Rule 3 reinforced Rule 2 by also showing that customers who bought DM safety products also purchased tools and storage products (with a 50% confidence level and 1.25 lift ratio). Rule 4 demonstrated that customers who purchased electrical supplies also bought safety supplies (with a 62.5% confidence level and 1.04 lift ratio). The results from the foregoing association rules offer insight for the host salesperson in terms of identifying prospective cross-selling opportunities.

#### SALESPERSON IMPLICATIONS OF CASE STUDY FINDINGS

The primary objective of this paper was to introduce three relatively useful data mining tools that can assist sales personnel to improve their efficiency and effectiveness in their sales territory. Through data mining of one industrial distributor salesperson's sales for a 12-month period, valuable insight into the host salesperson's target customer groups and target

**Table 5. Transformed Purchasing Data for Selected Customers**

Customer	LAMPS	BALLASTS	FIXTURES	PORTABLE LIGHTING	FLASHLIGHTS/BATTERY	ELECTRICAL SUPPLIES	DISTRIBUTION/CONTROL
1	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	1	0
4	1	0	0	0	0	1	0
5	0	0	0	0	0	0	0
6	0	0	1	0	0	0	1
7	0	0	1	0	0	1	1
8	1	0	0	0	0	1	0

**Note:** When a cell value equals "1", it indicates that the customer in that *row* purchased the product in that *column* in the past 12 months. If the cell value equals "0", it means that no such transactions could be found in the data set for that customer for the particular product.

**Table 6. Association Rules with Minimum Support of 4 and Minimum Consequence of 50%**

Rule #	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support(a U c)	Lift Ratio
1	80	A1-LAMPS=>	J1-SAFETY	5	12	4	1.333333
2	75	M1-TOOLS & STORAGE=>	J1-SAFETY	8	12	6	1.25
3	50	J1-SAFETY=>	M1-TOOLS & STORAGE	12	8	6	1.25
4	62.5	B1-ELECTRICAL SUPPLIES=>	J1-SAFETY	8	12	5	1.041667

products was obtained. Specifically, the findings of the three data mining techniques revealed information that affords this salesperson opportunity to design a sales plan tailored to her territory's customers. Purely as an illustration, actions that the host salesperson could take vis-à-vis the three data mining approaches are now offered.

#### **Sales Plan from Results of Multiple Linear Regression.**

Findings from the regression analysis revealed that the host salesperson should focus on four key variables in her efforts to grow sales in the territory. Specific actions that she might take are noted below.

1. Customers who are DM corporate accounts should be given enhanced discounts vis-à-vis local targeted accounts. The current discount program for which corporate accounts are eligible is having a negative impact on sales. Indeed, their discount program generates lower sales volume than discounts offered to local targeted customers, thus penalizing them for being corporate accounts. Therefore, the host salesperson should work with DM to develop an effective discount structure specifically for corporate accounts that will reverse the current adverse influence the present program has on their level of sales
2. Customers receiving pre-paid freight tend to

have much larger sales levels than do those who do not receive this concession. To receive pre-paid freight requires an account to have a sufficient level of sales to qualify for free shipping. Therefore, the host salesperson should work closely with those accounts not receiving free shipping (noted in Table 2) in efforts to enhance their sales to the qualifying level for pre-paid freight. Using the results of the market basket analysis (selling associated products) could also be utilized to help increase such accounts' sales.

3. Customers e-mailing or faxing their orders to DM have higher sales levels than accounts not doing so. Therefore, the host salesperson should work with those accounts not using either of the foregoing expedited ordering methods. Such customers need to be encouraged about the speed, convenience, and safety of submitting their orders electronically. In fact, DM might consider offering customers an incentive for sending their orders via e-mail or fax.

#### **Sales Plan from Results of Cluster Analysis**

Findings from the cluster analysis indicate that the sample territory can be divided into distinct clusters that should receive special selling attention. For example, the higher level cluster layer revealed an especially relevant segmentation outcome (Figure 1): Segment #1 with customers 1, 17, 6, 20, 5, 9, 12, 13, 14; Segment #2 with

customers 3, 15, 8, 11; and Segment #3 with customers 2, 7, 10, 19, 4, 18, and 16.

Segment #1 was comprised of manufacturing customers who spent a small amount of money on DM products relative to their MRO potential, were not corporate accounts, did not participate in the VMI program, and were not recipients of pre-paid freight. Clearly, these accounts merit special attention owing to their potential sales volume and lack of participation in DM programs. The host salesperson should make concerted efforts to enhance the sales from these accounts so that they can become eligible for DM customer programs and improved primary discounts.

Segment #2 consisted of health care institutions that were accorded corporate account status, whose 12-month sales essentially were low relative to their MRO potential, were not eligible for the VMI program, and had comparably few distinctly different buyers of DM products within their organization. Customers in this segment should be informed of the value of the VMI program, thus enticing them to increase their sales volume to qualify for it. Furthermore, the host salesperson should seek to cultivate additional buyers within each of these health care facilities; doing so could help augment sales from these customers.

Segment #3 constituted manufacturing accounts that generated a large sales volume for DM, may or may not be a corporate account, received pre-paid freight, may or may not have participated in the VMI program, and qualified for pre-paid freight. The host salesperson should be especially vigilant to protect the accounts in this segment from being enticed by the selling efforts of competitors. A loss of any of these customers would result in a substantial decline in sales revenue from the segment. Owing to the size of these accounts, special attention should be given

to these clients by providing extra customer service (e.g., ensuring accurate and timely deliveries, offering ways to optimize the use of DM products). Additionally, informing these customers about how they could receive higher discounts by augmenting their purchases from DM should be emphasized.

### **Sales Plan from Results of Association Rules**

Transaction information utilized for the study included every product category sold to the salesperson's twenty customers during a 12-month period. The association rules revealed particular propensities for customers who buy items in one product category to also buy items in another category. The rules with the highest number of support occurrences and lift ratios greater than 1 are illustrated in Table 6 and offer specific actions the host salesperson could take with her customers.

Findings indicate that customers purchasing tools and storage from DM also tend to buy safety products (Table 6). Indeed, DM is one of the largest distributors in North America of both of these product categories. Although tools are not closely related to safety products, safety products are DM's largest product category. Accordingly, a sales plan directed at selling these two product categories to any customer who is not purchasing from both of these categories seems warranted. The sales effort could entail use of targeted e-mail promotions, product demonstrations at the customer site, discount promotions, and cost savings demonstrations for high quality, economical, private-labeled, globally-sourced brands.

Shown in Table 6 is another set of association rule results. Findings from this analysis suggest that two other product categories which show associations with safety products are lamps and electrical supplies. This result suggests that a

similar sales plan as that above should be designed to present cross-selling opportunities to buyers of safety products and lamps and/or between safety products and electrical supplies.

Utilization of association rules analysis offers the host salesperson insight into the particular products that could be emphasized to grow sales. Concomitantly, generating increased sales from customers not only enhances the performance of the salesperson, but it also assists customers to generate a sales volume that affords qualification for prepaid freight and a deeper primary discount with DM (i.e., special concessions DM offers to qualifying customers).

Association rule results also imply that the host salesperson demonstrate to DM retail stores where to stock items. More specifically, the DM salesperson should convince DM stores in her territory to arrange items in close proximity to each other that are often purchased together (e.g., position safety products near the tools, lamps, and electrical supplies). Furthermore, an online selling effort could be directed at DM retail store customers by sending targeted e-mails from DM's web site. For example, the e-mail could state (à la Amazon.com) "customers who purchased this item also purchased the items noted below."

## MANAGERIAL CONCLUSIONS

Integrating data mining into a sales plan creates direction and focus with increased sales results and top-flight customer relationship management as the end goals. The sheer number of customer profiles, products, programs, and market segments can be overwhelming to a salesperson. Through the use of data mining, however, specific information can be uncovered that salespeople can utilize in the management of their sales territory. Such insights can assist sales personnel in maintaining

superior customer relationships. By applying data mining techniques to existing customer data, the potential exists for a salesperson to leverage the results to advantage. Findings from our case study suggest that the use of multiple linear regression, association rules, and cluster analysis as data mining techniques offer sales personnel customer-oriented tools that could enhance the consequences of their territory efforts. Two caveats about data mining, though, should be kept in mind. First, the findings obtained in a given salesperson's territory should not be generalized to other salespeople's territories; the results are likely to be territory specific. Therefore, sales personnel interested in using data mining should undertake their own efforts rather than relying on data mining knowledge gleaned from another salesperson. Second, statistics and analytics cannot replace human intuition. Salespeople must still carefully analyze the findings to assure that the data mining results are logical.

This paper has provided guidelines for utilizing data mining in the industrial supply industry. Using a sample sales territory, a framework was presented about how to apply three data mining techniques with actual sales data to generate feasible sales plans. In order to fully incorporate advantages of data mining into the sales force, periodic investigation of customer data is requisite. Once details of a sales plan are complete, the program should be launched. After a given period of time, data from the results of the new sales plan should be analyzed to test the effectiveness of the plan. The response data from that selling effort could be data mined, thus affording a new look at the data and potential modification of the sales plan.

Admittedly, many sales personnel are unlikely to be immediately skilled to undertake data mining on their own. After all, they generally are trained

to serve as conduits between their company and their accounts, not as statistical or computer technicians. Therefore, most salespeople interested in using data mining to enhance their relationships with each customer will need to become cognizant of (a) the importance of data mining, (b) useful data mining methods vis-à-vis their respective territories, (c) means of conducting these analyses and interpreting their results, and (d) requisite data to undertake data mining. As Hair et al. (2009, p. 258) state, "... sales reps need to know how to use the software, what information they are required to maintain electronically about their customers...and how to use equipment and applications...that enhance the automated sales environment." For progressive sales managers who see the wisdom of using data mining at a sales territory level, the following recommendations are offered:

1. In initial and refresher sales training programs, trainees should be made cognizant of what is meant by data mining and how its tools can be beneficial to sales personnel wishing to make use of it. Lectures may be ideal for dissemination of such information. Firms that already use data mining, but at a macro (company) level, should present to trainees the value derived from utilizing such efforts. Company employees who use data mining should serve as trainers. After all, they are conversant with data mining's strengths, weaknesses, and requisite knowledge, particularly vis-à-vis the firm and its industry.
2. Sales personnel should be trained about what data are necessary to use various data mining tools. This training could be received at a community college, a four-year university, an online university, or the company itself. The salespeople's firm should inform salespersons what data are readily available in

its data base for undertaking data mining efforts.

3. Company salespeople will need to know how to actually conduct various data mining techniques (such as those used by the host company salesperson in our case study). Again, this training could be received at a community college, a four-year university, an online university, or the company. Firms that make extensive use of data mining at a company-wide level could have its data mining experts do the training, as they can relate specifically to the environment in which its company's sales personnel interact.
4. Salespeople whose territories are extremely small (e.g., fewer than 10 customers) should be apprised by company personnel that use of data mining might not be particularly beneficial. Such sales personnel may well find that acquiring unique knowledge about each of their customers through the use of in-depth interviews may yield more reliable results than relying on data mining tools.
5. To enhance understanding of data mining techniques and interpretations of those techniques, management may suggest to salespersons that they enroll in a basic statistics course at some college-level institution to facilitate their use of data mining tools.

## END NOTES

- 1 Classification analysis was not conducted because all customer types were known to the host salesperson.
- 2 Because of the confidential nature of customer records, data were only available from the host salesperson's territory.
- 3 Table 2 is presented to provide readers with a foretaste of products the host salesperson's



customers purchase. As such, information about all twenty customers is not shown. The brevity is consistent with DM management's request to maintain as much confidentiality as possible.

- 4 For a detailed discussion about how to interpret multiple linear regression results, see Hair et al. (2010). Although multiple linear regression is often employed to suggest a cause-effect relationship between two variables, caution should be exercised regarding the multiple linear regression findings in our case study. In the present analysis, the significant independent variables are correlated with customer sales revenue; whether they definitively lead to a higher (or lower) level of sales, however, is uncertain.
- 5 For detailed discussion about how to conduct cluster analysis, see Hair et al. (2010). Prototypical software packages containing cluster analysis include SAS, SPSS, STATA, and XLMiner for Excel. Each software package provides essentially similar options for cluster analysis.
- 6 For detailed discussion about how to conduct association rules, see Schmueli et al. (2007) and Shaw et al. (2001). Prototypical software packages containing cluster analysis include SPSS Clementine, IBM Intelligent Miner for Data and XLMiner for Excel. Each software package provides essentially similar options for association rules.
- 7 In addition, irregular purchases had to be eliminated from the dataset so that the finding would reflect the customer's regular purchasing patterns. According to Kumar (2002), a cut-off rule has to be identified to rule out the "accidental" purchase which might lead to an "accidental" association.

*Northern Illinois University*

Therefore, different thresholds for eliminating the irregular purchase (also called a "spot buy") were tested. After these tests, the host salesperson discerned that if a purchase value was less than 5% of the customer's total annual sales, it would be regarded as an irregular purchase, and thus, that particular purchase would be eliminated when conducting the binary (0,1) data transformation.

---

Alan J. Dubinsky is the Dillard Distinguished Professor of Marketing at Dillard College of Business Administration, Midwestern State University in Wichita Falls, TX. He is also Professor Emeritus at Purdue University in West Lafayette, IN. Email: [Dubinsky@purdue.edu](mailto:Dubinsky@purdue.edu)

Guoying Zhang is an Assistant Professor of Management Information Systems in the Dillard College of Business Administration, Midwestern State University in Wichita Falls, TX. Email: [grace.zhang@mwsu.edu](mailto:grace.zhang@mwsu.edu)

Michele Wood, MBA, Dillard College of Business Administration at Midwestern State University in Wichita Falls, TX. Email: [Michele\\_wood@sbcglobal.net](mailto:Michele_wood@sbcglobal.net)

## REFERENCES

- Anderson, R. E., A.J. Dubinsky, and R. Mehta (2007), *Personal Selling: Building Customer Relationships and Partnerships*, Florence, KY: South-Western/Cengage Learning.
- Berry, M.J. and G. Linoff (1997), *Data Mining Techniques for Marketing, Sales, and Customer Support*, Hoboken, NJ: John Wiley & Sons, Inc.
- Berson, A., S. Smith, and K. Thearling (1999), *Building Data Mining Applications for CRM*, New York, NY: McGraw-Hill Companies.

- Boulding, W., R. Staelin, M. Ehret, and W.J. Johnston (2005), "A CRM Roadmap: What We Know, Potential Pitfalls, and Where to Go," *Journal of Marketing*, 69 (October), 155-167.
- Cao, Y. and T.S. Gruca (2005), "Reducing Adverse Selection through Customer Relationship Management," *Journal of Marketing*, 69 (October), 219-229.
- Chen, Y., G. Zhang, D. Hu, and S. Wang (2006), "Customer Segmentation in Customer Relationship Management Based on Data Mining," *Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management*, 207, 288-293.
- Chou, A., E.B. Pullins, and S. Senecal (2007), "Technology Empowerment As a Determinant of Salesforce Technology Usage," *Journal of Selling and Major Account Management*, 7 (2), 20-29.
- Company DM (2008), *Company DM 2008 Annual Report*.
- Crosby, L.A. and S.L. Johnson (2001), "High Performance Marketing in the CRM Era," *Marketing Management*, 10, (3), 10-11.
- Dickson, P.R., W.M. Lassar, G. Hunter, and S. Chakravorti (2009), "The Pursuit of Excellence in Process Thinking and Customer Relationship Management," *Journal of Personal Selling and Sales Management*, 24 (Spring), 111-124.
- Dubinsky, A.J. (1999), "When Salespeople Fail: Assessing Blame," *Industrial Marketing Management*, 28 (January), 19-26.
- Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth (1996), "From Data Mining to Knowledge Discovery in Databases," *AI Magazine* (Fall), 37-54.
- Hair, J.F., W.C. Black, B.J. Babin, and R.E. Anderson (2010), *Multivariate Data Analysis*, Upper Saddle River, NJ: Prentice Hall.
- Hair, J.F., R.E. Anderson, R. Mehta, and B.J. Babin (2009), *Sales Management*, Boston: Houghton Mifflin Company.
- Hunter, G.K. and W.D. Perreault (2007), "Making Sales Technology Effective," *Journal of Marketing*, 71 (1), 16-34.
- Javalgi, R.G., C.L. Martin, and R.B. Young (2006), "Market Research, Market Orientation, and Customer Relationship Management: A Framework and Implications for Service Providers," *Journal of Services Marketing*, 20 (1), 12-23.
- Kale, S.H. (2004), "CRM Failure and the Seven Deadly Sins," *Marketing Management*, 13 (#5), 42-46.
- Landry, T.D., T.J. Arnold, and A. Arndt (2005), "A Compendium of Sales-Related Literature in Customer Relationship Management: Processes and Technologies with Managerial Implications," *Journal of Personal Selling and Sales Management*, 25 (Summer), 231-251.
- Mathieu, J., M. Ahearne, and S.R. Taylor (2007), "A Longitudinal Cross-Level Model of Leader and Salesperson Influences on Sales Force Technology Use and Performance," *Journal of Applied Psychology*, 92 (2), 528-537.
- Moutot, J.M. and G. Bascoul (2008), "Effects of Sales Force Automation on Sales Force Activities and Customer Relationship Management Processes," *Journal of Personal Selling and Sales Management*, 28 (Spring), 167-184.
- Olaffson, S., X. Li, and S. Wu (2006), "Operations Research and Data Mining," *European Journal of Operational Research*, 187, 1429-1448.
- Payne, A. and P. Frow (2006), "Customer Relationship Management: From Strategy to Implementation," *Journal of Marketing Management*, 22, 135-168.

- Payne, A. and P. Frow (2005), "A Strategic Framework for Customer Relationship Management," *Journal of Marketing*, 69 (October), 167-176.
- Raman, P., C.M. Wittmann, and N.A. Rauseo (2006), "Leveraging CRM for Sales: The Role of Organizational Capabilities in Successful CRM Implementation," *Journal of Personal Selling and Sales Management*, 26 (Winter), 39-53.
- Reinartz, W., M. Krafft, W.D. Hoyer (2004), "The Customer Relationship Management Process: Its Measurement and Impact on Performance," *Journal of Marketing Research*, 41 (August), 293-305.
- Sharma, A. and J.N. Sheth (2010), "A Framework of Technology Mediation in Consumer Selling: Implications for Firms and Sales Management," *Journal of Personal Selling and Sales Management*, 30 (Spring), 121-129.
- Shaw, M.J., C. Subramaniam, G.W. Tan, and M.E. Welge (2001), "Knowledge Management and Data Mining for Marketing," *Decision Support Systems*, 31, 127-137.
- Shmueli, G., N.R. Patel, and P.C. Bruce. (2007), *Data Mining for Business Intelligence*, Hoboken, NJ: John Wiley & Sons, Inc.
- Stein, A. and M. Smith (2009), "CRM Systems and Organizational Learning: An Exploration of the Relationship between CRM Effectiveness and the Customer Information Orientation of the Firm in Industrial Markets," *Industrial Marketing Management*, 38, 198-206.
- Stoddard, J.E., S.W. Clopton, and R.A. Avila (2006), "An Analysis of the Effects of Sales Force Automation on Salesperson Perceptions of Performance," *Journal of Selling and Major Account Management*, 6 (1), 38-56.
- Sumathi, S. and S. N. Sivanandam (2006), "Data Mining in Customer Value and Customer Relationship Management," *Studies in Computational Intelligence*, 29, 321-386.
- Tan P., V. Kumar, and J. Srivastava (2002), "Selecting the Right Interestingness Measure for Associations Patterns," *Proceedings of the 8th ACM SIGKDD International Conference on KDD*, 32-41.
- Tu, Y. (2008), "An Application of Web-Based Data Mining: Selling Strategies for Online Auctions," *Online Information Review*, 32 (2), 147-162.
- Yim, F.H.K. (2002), "CRM Orientation: Conceptualization and Scale Development," Master's Thesis, Shatin: Chinese University of Hong Kong.
- Yim, F.H.K., R.E. Anderson, and S. Swaminathan (2004), "Customer Relationship Management: Its Dimensions and Effect on Customer Outcomes," *Journal of Personal Selling & Sales Management*, 24 (Fall), 263-278.
- Zhang, G., F. Zhou, F. Wang, and J. Luo (2008), "Knowledge Creation in Marketing Based on Data Mining," in *2008 International Conference on Intelligent Computation Technology and Automation*, Hunan, China: ICICTA, 782-786.