

IBM Research Report

Segmentation-Based Modeling for Advanced Targeted Marketing

C. Apte, E. Bibelnieks, R. Natarajan, E. Pednault, F. Tipu

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

D. Campbell, B. Nelson
Fingerhut Business Intelligence
4400 Baker Road
Minnetonka, MN 55343



Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

Segmentation-Based Modeling for Advanced Targeted Marketing¹

C. Apte, E. Bibelnieks, R. Natarajan, E. Pednault, F. Tipu

T.J. Watson Research Center, IBM, P.O. Box 218, Yorktown Heights, NY 10598

D. Campbell, B. Nelson

Fingerhut Business Intelligence, 4400 Baker Road, Minnetonka, MN 55343

Abstract

Fingerhut Business Intelligence (BI) has a long and successful history of building statistical models to predict consumer behavior, and it constantly strives to improve its decision-making processes and tools. Fingerhut has found that predictive models can be much more effective when the target audience is split into subpopulations (i.e., customer segments) and individually tailored predictive models are developed for each segment. Historically, Fingerhut BI has used decision trees or simply domain expertise for creating customer segments. Even though these approaches work well, they are “sub-optimal” because effectiveness (i.e., predictive strength) of the segment models is not considered when defining the segments. Given their mailing volumes, Fingerhut is sensitive to the fact that increasing the predictive power of their models means millions of dollars in new revenue. Fingerhut BI approached IBM Research with the problem of how to build segmentation-based models more effectively so as to maximize predictive accuracy. The IBM Advanced Targeted Marketing – Single Events™ (IBM ATM-SE™) solution is the result of IBM Research and Fingerhut BI directing their efforts jointly towards solving this problem. This paper presents an evaluation of ATM-SE’s modeling capabilities using data from Fingerhut’s catalog mailings.

¹ Contact information: Chid Apte, IBM, Route 134 at Taconic Parkway, P.O. Box 218, Yorktown Heights, NY 10598. e-mail: apte@us.ibm.com phone: 914-945-1024 fax: 914-945-3434

1 BACKGROUND

Many direct marketers or targeted marketers build numerous models each year to predict customer responses to their offers. These offers include customer acquisition, up-sell, cross-sell, reactivation, and others aimed at building customer relationships. The customers eligible for these offers include new consumers who have never purchased before, one-time buyers, multi-buyers, and inactive buyers. The data used for these models typically include both demographic and behavioral information. As a rule, the most predictive data is behavioral information, such as historical purchase data that targeted marketers capture through their campaign management systems.

Catalog offers are constructed with a target audience in mind. Analysis of customer data is then performed to identify the best opportunities within the target audience. For example, at Fingerhut, a Spring Electronics catalog may be targeted to customers who have exhibited a preference for stereos, computers, CD players, televisions and cameras, and purchase in the spring. Similarly, a Winter Outdoor Catalog may be targeted to customers who participate in winter sports such as cross-country skiing, snow boarding, snowshoeing, and winter camping. Fingerhut could select customers for each mailing using customer segmentation criteria traditionally employed in the targeted marketing industry, such as recency (days since last purchase), frequency (number of purchases per year), and monetary value (dollars spent over the last n years) in the specified product categories. However, the company has found that statistical modeling techniques based on hundreds of customer-level data fields are superior for selecting customers that are most likely to buy from their catalog offers.

Fingerhut Business Intelligence (BI) has a long and successful history of building statistical models to predict consumer behavior, and it constantly strives to improve its decision-making processes and tools [1, 3, 5, 7, 13, 14]. Based on this expertise, Fingerhut has found that predictive models can be much more effective when the target audience is split into subpopulations (i.e., customer segments) and individually tailored predictive models are developed for each segment. Historically, Fingerhut BI has used decision trees or simply domain expertise for creating customer segments. Even though these approaches work well, they are “sub-optimal” because effectiveness (i.e., predictive strength) of the segment models is not considered when defining the segments.

Given their mailing volumes, Fingerhut is sensitive to the fact that increasing the predictive power of their models means millions of dollars in new revenue. Fingerhut BI approached IBM Research with the problem of how to build segmentation-based models more effectively so as to maximize predictive accuracy. The IBM Advanced Targeted Marketing – Single Events™ (IBM ATM-SE™) solution is the

result of IBM Research and Fingerhut BI directing their efforts jointly towards solving this problem. That is, ATM-SE was developed to meet the segmentation-based modeling needs of targeted marketers.

2 SOLUTION CAPABILITIES

The ATM-SE solution is a client-server application. The user interface client allows the analyst to manage the data mining tasks performed by the ATM-SE server and to view the results. Actual data mining is performed by the underlying data-mining engine, which is invoked by the ATM-SE server. The ATM-SE server and client may be running on the same machine or on different machines connected by a network. The user has the ability to inspect the meta-data associated with the data sets available for mining, set up data mining runs, examine results from completed data mining runs using rule or tree visualizations, evaluate models on test data using lift charts or statistical measures, and export the models as SAS or C code for operational deployment. The ATM-SE server utilizes the IBM ProbE™ data mining engine for performing two predictive modeling tasks. One algorithmic approach is used in predicting continuous outcomes, such as expected sales revenue for each individual customer. The second approach is used in predicting categorical outcomes, such as customer response (i.e., whether or not each individual customer will respond to a mailing).

ProbE (pronounced *probe*, for *Probabilistic Estimation*) is a customizable data mining engine that is being developed to enhance the commercial state-of-the-art in predictive modeling products and services. ProbE might best be described as an extensible, embeddable, and scalable segmentation-based modeling engine. The design of ProbE has been motivated by recent advances in integrating statistics and learning techniques with data management [6, 8, 11]. ProbE's application programming interfaces (API's) are particularly well-suited for implementing segmentation-based modeling techniques, wherein data records are partitioned into segments and separate predictive models are developed for each segment.

At the time the ATM-SE solution was conceived, Fingerhut BI was already employing segmentation-based modeling methodologies. This style of modeling is popular among data analysts and applied statisticians in general. However, as previously discussed, it is usually approached as a sequential process in which data is first segmented (using, for example, conventional decision tree algorithms, unsupervised clustering algorithms, domain expertise, intuition, etc.) and predictive models are then developed for those segments. The drawback of this sequential approach is that it ignores the strong influence that segmentation exerts on the predictive accuracies of the models within each segment

ProbE, on the other hand, is able to perform segmentation and predictive modeling within each segment simultaneously, thereby optimizing the segmentation so as to maximize overall predictive accuracy. The

benefit of this optimizing approach is that it can produce better models than might otherwise be obtained. ProbE has thus far been found to consistently produce high-quality models on a fully automated basis without requiring costly manual adjustments of the models or the mining parameters by data mining experts. The latter property is mandatory in order to make data mining attractive to medium-sized businesses.

Another key feature of ProbE is that it can be readily extended so as to construct virtually any kind of predictive model within a segment. For example, in the case of the ATM-SE solution, least-squares linear regression with forward stepwise feature selection is used to construct segment models for continuous target variables. For categorical target variables, naï ve Bayes modeling with forward stepwise feature selection is employed. The detailed technical descriptions of these algorithms appear in separate papers [9, 10]. In the case of the IBM Underwriting Profitability Analysis™ (IBM UPA™) application [2], joint Poisson/Log-Normal statistical models are constructed for each segment to simultaneously model both the frequency with which insurance claims are filed and the amounts (i.e., severities) of those claims.

The segmentations that are produced depend strongly on the nature of the predictive models employed in each segment. For example, when stepwise linear regression is used, the resulting segments correspond to regions of the response surface that are locally linear and the boundaries between segments correspond to non-linearities detected in the response surface. In the case of joint Poisson/log-normal models, the resulting segments correspond to distinct risk groups whose loss characteristics (i.e., frequency and severity) are estimated in accordance with standard actuarial practices.

3 EVALUATION

In order to evaluate the ATM-SE solution, analysts at Fingerhut Business Intelligence used the solution to build several different segmentation-based models utilizing both the stepwise linear regression and stepwise naï ve Bayes capabilities. All models were developed using training data from a Fingerhut GMC (General Merchandise Catalog) mailing from the fall of 1998. The resulting models were then evaluated using data from two GMC mailings from the fall of 1999. Existing GMC models were used as the benchmark or comparison models. Both the ATM-SE models and the benchmark models were developed using a suite of over 1,400 customer-level predictor variables commonly employed by Fingerhut in their modeling efforts. The training set contained about half a million records, while the two validation sets contained about a quarter million records each.

The ATM-SE models and the corresponding benchmark models were compared in terms of their segmentation power (i.e., lift charts) and accuracy (i.e., predictions versus actual outcomes). No

comparisons were made to models produced using other predictive modeling technologies. Over the years, Fingerhut Business Intelligence has evaluated numerous data mining and statistical modeling technologies. Fingerhut’s proprietary benchmark models have almost always surpassed predictive models constructed using other technologies. The benchmark models therefore represent the state of the art in terms of what can be achieved using existing modeling methodologies.

3.1 Summary of Results

For the evaluation study, Fingerhut constructed four models using the ATM-SE solution. The models are summarized in Table 1. In the case of Models A and D, the target variable was sales revenue. The training data for Model A included both buyers and non-buyers (i.e., non-responders whose sales revenues were therefore zero). Model A therefore predicts expected sales revenue. The training data for Model D, on the other hand, included buyers only. Model D therefore predicts expected sales revenue given that a customer responds to the mailing. In the case of Models B and C, the target variable was a response indicator with a value of 1 for buyers, and 0 for non-buyers. These models therefore predict the probability of response.

Lift curves were constructed for each of the models in Table 1 using data from the second of the two validation mailings used in the study. Table 2 reports the areas under the lift curves that were computed for the ATM-SE models and the corresponding benchmark models. Generally speaking, the greater the area, the better the model. As can be seen from Table 2, the ATM-SE models consistently outperformed the corresponding benchmark models in terms of the areas under their lift curves. Model B*D in Table 2 represents the joint model produced by multiplying the outputs of Model B (probability of response) and Model D (expected sales revenue given response) in order to predict (unconditional) expected sales revenue. Fingerhut bases its mailing decisions on expected sales revenue together with other considerations, such as credit worthiness and mail saturation effects [3, 5, 13].

Model ID	Segment Model Type	Output of Resulting Model
A	Stepwise Linear Regression	Expected Sales Revenue
B	Stepwise Naï ve Bayes	Probability of Response
C	Stepwise Linear Regression	Probability of Response
D	Stepwise Linear Regression	Expected Sales Revenue Given Response

Table 1. The segment model types and the model outputs of the four models constructed by Fingerhut BI using the ATM-SE solution.

Model ID	ATM-SE Model	Benchmark Model
A	64.3	63.9
B	60.9	60.1
C	60.7	60.1
B*D	64.1	63.9

Table 2. Areas under the lift curves for the ATM-SE models and the corresponding benchmark models for Validation Mailing 2. In each comparison, the larger of the two areas is highlighted in bold.

Figure 1 shows the lift curves for ATM-SE Models A and B*D using data from Validation Mailing 2. Figure 1 also shows the lift curve for the corresponding benchmark model. As can be seen from these lift curves and Table 1, Model A produces the best overall lift. Model B*D produces slightly lower lift for high-ranking customers (i.e., those at the lower end of the X-axis). However, the lift of Model B*D is slightly higher than that of Model A for low-ranking customers (i.e., those at the higher end of the X-axis). The lift of the benchmark model closely matches that of Model A for high-ranking customers; however, Models A and B*D both produce higher lift than the benchmark model for low-ranking customers. Models A and B*D are thus better able to separate low-ranking, low-profit customers from the rest of the population.

As reported below, ATM-SE Model A likewise outperformed the corresponding benchmark model on the first of the two validation mailings used in the study. In particular, the same relationships observed in Figure 1 between Model A and the corresponding benchmark model are observed on Validation Mailing 1. Analyses of Models B, C and D were not performed on Validation Mailing 1 because of time constraints during the evaluation study. Due to space limitations, only detailed evaluations of Model A are presented in this paper.

3.2 Model A

ATM-SE Model A consists of eight mutually exclusive segments, each of which has a corresponding linear regression equation to predict expected sales revenue. Figure 2 shows the decision tree that defines the eight segments. The variables that appear in the regression equations for each segment are listed in Table 3. Note that the Probe data mining engine used in the ATM-SE solution performs feature selection not only when building trees, but also when building segment models. Moreover, the variables selected for the segment models need not be the same as the variables that define the segments. This degree of flexibility is necessary to maximize predictive accuracy.

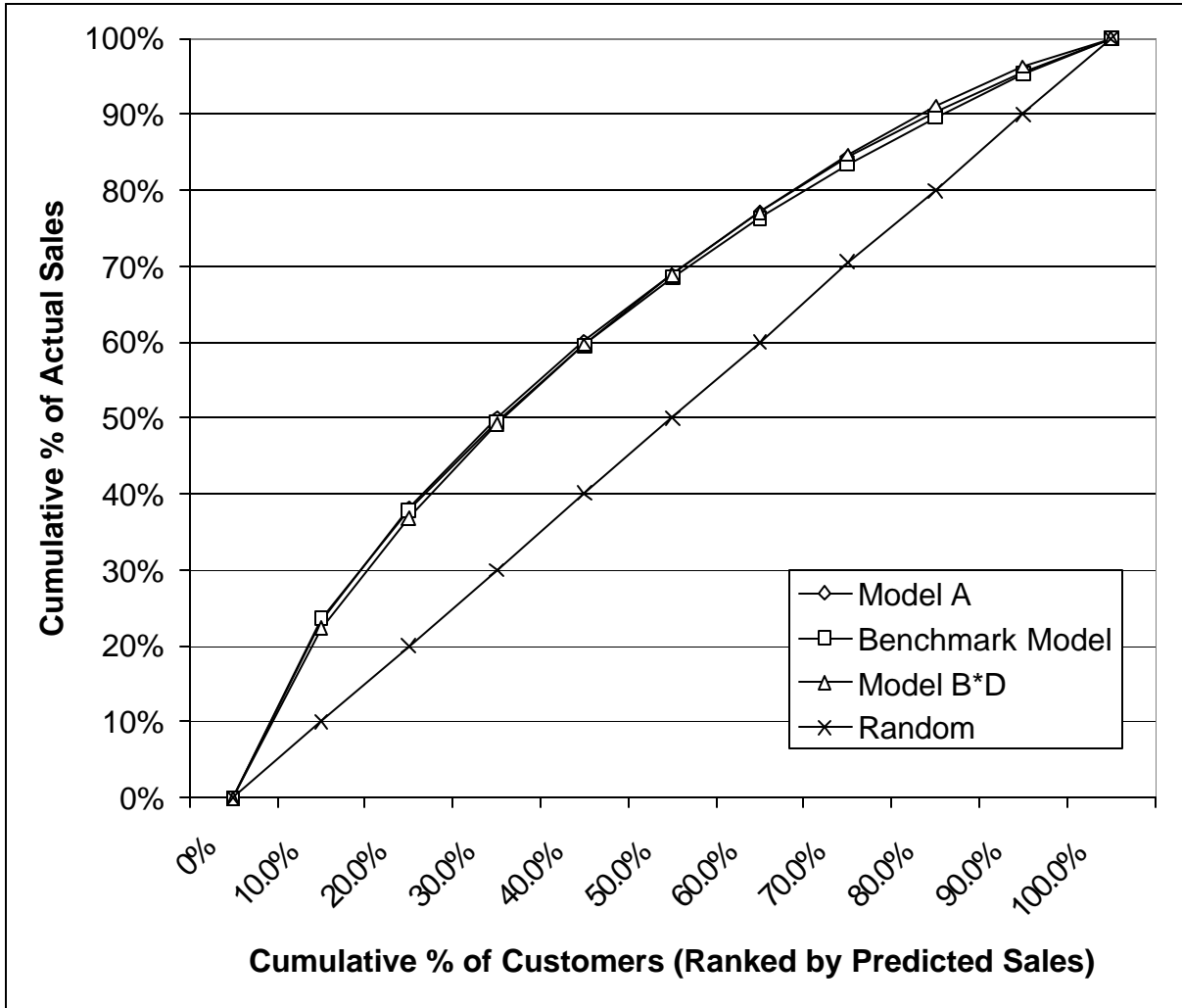


Figure 1. Power of segmentation curves for ATM-SE Models A and B*D, and the corresponding benchmark model, using data from Validation Mailing 2.

To perform the evaluation, Fingerhut Business Intelligence coded both the segment definitions and the corresponding regression models in SAS. At the time ModelA was constructed, the ATM-SE solution could output models only in text form, which were then manually translated into SAS code. Since this model was built, ATM-SE has been enhanced to output SAS code directly. This latter feature was used to generate SAS code to evaluate Model B*D as reported above.

Table 4 shows the areas under the lift curves obtained for both ModelA and the corresponding benchmark model on both validation mailings. Note that the entries for Validation Mailing 2 are the same as those in Table 2. The benchmark model for Model A is actually a combination of two models: one that

predicts the probability of response; the other that predicts expected sales revenue given a response. The two individual model outputs are multiplied together to predict (unconditional) expected sales revenue. In the case of the ATM-SE model, expected sales revenue is predicted directly.

Two other accuracy measures likewise imply that ATM-SE Model A is equal to or more accurate than the corresponding benchmark model. Table 5 shows the Pearson correlation coefficients between the actual sales revenues and the predicted sales revenues of the two models for the two validation mailings. For both validation mailings, the correlation of predicted and actual sales is the same or higher for the ATM-SE model as compared to the benchmark model. Table 6 shows the mean absolute differences for the two models (i.e., the sum of the absolute values of the differences between actual and predicted sales revenues, divided by the number of data records). In equation form, the mean absolute difference is given

by
$$MAD = \frac{1}{n} \sum_{i=1}^n |predicted_i - actual_i|$$
. For both mailings, the MAD statistic is lower for the ATM-SE

model than for the benchmark model, implying that the ATM-SE model contains less error or is more accurate than the benchmark model.

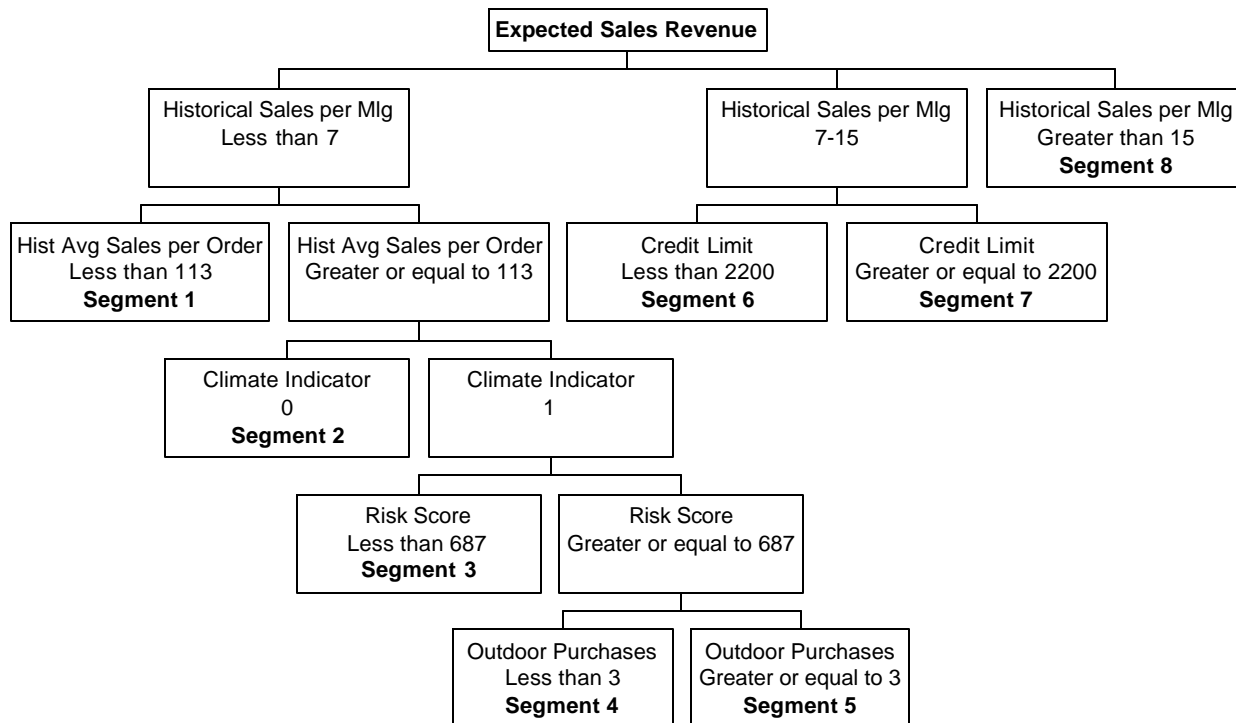


Figure 2. The decision tree that defines the population segments of ATM-SE ModelA.

Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8
Hist \$/Order Hist # Orders	<i>constant</i>	Hist Dollars Hist # Orders	Ave \$/Mlg Hist # Orders	<i>constant</i>	Hist # Orders Cancel \$ Phone Orders	Hist Dollars Hist # Orders	Hist # Orders Ave \$/Mlg Hist Dollars Cancel \$

Table 3. The variables used in each segment model of ModelA.

Validation Mailing	ATM-SE Model A	Benchmark Model
1	67.3	66.9
2	64.3	63.9

Table 4. Areas under the lift curves for ModelA and the corresponding benchmark model. For each validation mailing, the larger of the two areas is highlighted in bold.

Validation Mailing	ATM-SE Model A	Benchmark Model
1	.10	.10
2	.078	.076

Table 5. Pearson product moments for ModelA and the corresponding benchmark model. For each validation mailing, the larger of the two correlations is highlighted in bold.

Validation Mailing	ATM-SE Model A	Benchmark Model
1	23.29	25.38
2	19.10	21.02

Table 6. Mean absolute differences for ModelA and the corresponding benchmark model. For each validation mailing, the smaller of the two error measures is highlighted in bold.

4 CONCLUSIONS

The basic philosophy behind the IBM Probe data mining engine is to augment the automatic segmentation capability of decision tree algorithms [4, 12] and other segmentation algorithms with the use of more elaborate statistical models in each segment. Our initial work considered joint Poisson/log-normal models for insurance risk modeling. This work resulted in the IBM UPA application. Our most recent work has focused on segment models that incorporate local covariate dependencies and feature selection (i.e., stepwise linear regression and stepwise naï ve Bayes). We have observed that this latter approach produces models that are better adapted to the structure of the response data and its covariate

dependencies, leading to improved interpretability and predictive accuracy over most existing modeling methods.

Significant advances had to be made to Probe's segmentation algorithms as initially developed for the UPA application in order to allow segment models to incorporate local covariate dependencies and feature selection. Preliminary work on using stepwise linear regression for segment models led to the joint project with Fingerhut that produced the ATM-SE solution. During the course of the project, further enhancements had to be made to Probe's segmentation algorithms to construct naïve Bayes models in each segment and to refine the stepwise linear regression capability.

As a result of initial testing, Fingerhut Business Intelligence finds that Probe's ability to combine segmentation with stepwise linear regression and stepwise naïve Bayes works quite well. Even though Fingerhut's proprietary benchmark models typically outdo other techniques used in the data mining and statistical modeling arenas, all of the ATM-SE models were found to meet or slightly beat these benchmark models in terms of segmentation power and predictive accuracy. It is also important to note that this level of performance was achieved using the ATM-SE solution in a completely automated mode of operation, with no intervention from an experienced modeler.

As the analysts at Fingerhut become more familiar with the ATM-SE solution and incorporate their own expertise into the modeling process, the benefits can only increase. Because, the ATM-SE solution alleviates the burden of feature selection and model development, more attention can be devoted to data transformation and feature construction. The latter could well yield further improvements in model accuracy. In addition, as the understanding and experience with the solution grows, it is expected that the ATM-SE solution will be applied to a wider range of modeling applications at Fingerhut.

In a more general context, the Advanced Targeted Marketing – Single Events (ATM-SE) solution was designed to meet the advanced modeling needs of targeted marketers. With ATM-SE, customer segments and the corresponding models for those segments are constructed simultaneously in order to maximize the overall predictive accuracy of the resulting model. This approach is capable of producing higher-quality models than techniques in which segmentation and modeling are performed sequentially. In particular, the models produced by ATM-SE should provide significant improvements over models constructed using traditional recency-frequency-monetary segmentation methodologies employed by targeted marketers.

REFERENCES

1. Anthes, G., *Optimal Results*. Computerworld, 2000. **34**(47): p. 60-61.
2. Apte, C., et al., *Probabilistic Estimation Based Data Mining for Discovering Insurance Risks*. IEEE Intelligent Systems, 1999. **14**(6): p. 49-58.
3. Bibelnieks, E. and D. Campbell, *Mail Stream Streamlining*. Catalog Age, 2000. **17**(12): p. 118-120.
4. Breiman, L., et al., *Classification and Regression Trees*. 1984, Monterrey, CA.: Wadsworth.
5. Campbell, D., et al., *Optimizing Customer Mail Streams at Fingerhut*. Interfaces, 2001. To appear.
6. Hosking, J., E. Pednault, and M. Sudan, *A Statistical Perspective on Data Mining*. Future Generation Computer Systems, 1997. **13**(2-3): p. 117-134.
7. Lachs, J., *Data Mining Digs In*. American Demographics, 1999. **21**(7): p. 38-45.
8. Matheus, C.J., P.K. Chan, and G. Piatetsky-Shapiro, *Systems for Knowledge Discovery in Databases*. IEEE Transactions on Knowledge and Data Engineering, 1993. **5**(6): p. 903-913.
9. Natarajan, R. and E.P.D. Pednault, *Decision Trees with Node Regression Estimators for Massive Data Sets*. 2001. Submitted to ACM SIGKDD 2001.
10. Natarajan, R. and E.P.D. Pednault. *Using Simulated Pseudo Data To Speed Up Statistical Predictive Modeling From Massive Data Sets*. in *First SIAM International Conference on Data Mining*. 2001. Chicago. To appear.
11. Pednault, E.P.D., *Statistical Learning Theory*, in *The MIT Encyclopedia of the Cognitive Sciences*, R.A. Wilson and F.C. Keil, Editors. 1998, MIT Press: Cambridge, MA. p. 798-801.
12. Quinlan, J.R., *C4.5: Programs for Machine Learning*. 1993: Morgan Kaufmann.
13. Staff_Reporter, *Mail Order Giant Uses Business Intelligence to Make Every Mailing Count*. KM World, 1999. **8**(5).
14. Wreden, N., *Making Marketing Personal*. Beyond Computing, 1999: p. 24-25.