Communications of the IIMA

Volume 6 | Issue 1 Article 10

2006

Monitoring Sale Transactions for Illegal Activity

Robert J. Richardson
Iona College

Monitoring Sale Transactions for Illegal Activity

Robert J. Richardson
Hagan School of Business
Iona College, New Rochelle, NY 10801
(914)637-7725
rrichardson@iona.edu

ABSTRACT

Discriminant analysis and neural network methodologies were applied to the problem of identifying illegal sales transactions. The researchers independently developed models using data provided by a credit card company. A series of measures were developed and used to construct the models. The final results were that the discriminant analysis model recognized 32.3% of the fraudulent activity, while the neural network approach found 28.9%. With only 11.6% of the transactions in common, the combination of the two models identified 49.6%. In order to provide a real time monitoring program, the models were simplified yielding a capture rate of approximately 42%.

INTRODUCTION

The project for a major credit card encompassed two efforts to detect fraudulent patterns of cardholder's purchasing activity. The central focus of the investigation evaluated the use of statistical and neural network techniques. The final results are reported in this paper without disclosing specific details that would assist criminals in evading detection. This is consistent with other research that restricted the availability of data and censored results in their reports (e.g., Leonard, 1993).

Credit card operations expose the lending institution to two primary types of risk, credit and fraud. All circumstances where a cardholder or merchant become indebted to a bank without deception and is unable or unwilling to repay are classified as credit losses. All other situations are classified as fraud. Fraud is a crime although there are variations in its definition among the statutes of various countries where the credit card is used. Fraud includes the following categories: lost, stolen, not received, counterfeit, fraudulent application, fraudulent use of card, and other (Bolton and Hand, 2002). This study focuses on the fraudulent use of a stolen card or account number. In this situation, the thief spends as much as possible in as short a time as possible, before the theft is detected and the credit card is stopped. Therefore, the credit card company needs to detect the theft early to prevent large losses. For example, at the start of the study, the current detection methodology required that one hundred calls be made to cardholders for each stolen card detected. At a cost of $15 per call, the total cost to identify a fraud account is $1,500 which is equal the average loss saved by stopping the card. Management wanted earlier detection to reduce of the number of calls to identify each fraudulent card and an increase the amount saved by quicker action.

Specifically, the objectives of the project were to develop quantitative approaches to risk pattern analysis for use in real time processing and to develop procedures that significantly increase the bank's ability to identify and control risky transaction patterns. For the credit card industry, fraud accounts for over $850 million dollars in losses each year in the United States (Ghosh and Reilly, 1994) and $10 billion worldwide (Alerkerov, Freislebent and Rao, 1997). Although fraud losses are high in absolute dollars, they are only a small proportion of total activity. Thus, the problem of risk pattern recognition can be characterized as looking for a small number of needles in an enormously large haystack. These nonstandard transactions are identified as outliers that are detected by a number of different tests (Barnett and Lewis, 1994; Yu, Qian, Lu, and Zhou, 2006). Fraud is almost always directed toward cash or items that can be converted into cash and compressed in time. Thus, over time, the pattern of fraud use should diverge from legitimate use in recognizable ways.

The data used in this study include actual transactions organized by account over a six month period. Fraudulent transactions are identified. For example, a thief verifies that a stolen credit card is active by purchasing gasoline at a safe, automatic pump before going to a store (Provost, 2002). This allows us to segment the accounts into good and fraudulent transactions defined as those that had one or more fraudulent transactions. Over ten million transactions were used in the analysis.

RESEARCH ACTIVITIES

The research is directed at developing a methodology for identifying fraudulent transactions using a standardized scale to measure the probability that a given transaction is suspicious. The project plan includes the following steps: (1) develop measures, (2) select transactions for analysis, (3) use statistical techniques to develop models, (4) evaluate the accuracy of the models, and (5) implement the models for online process detection.

Measures

The final accuracy of the model depends on the measures or variables used to identify patterns in the sequence of transactions for each account. This is the creative step where all possible measures are generated. The initial measures are the source data for each transaction, i.e., the transaction amount. Next a series of measures are derived using the source measures, i.e., the change between consecutive transaction amounts. To gain a further prospective over time, moving averages are computed for the appropriate measures and become the basis for the final set of twenty-four measures.

Each of the source measures provides information on the current transaction, while the derived measures supply additional knowledge over a period of time. The change in consecutive purchase amounts involves the prior transaction and its relationship to the current transaction. The high-low range covers a longer time horizon that uses the last six transactions in its computation. To increase the effectiveness in determining patterns over time, the moving average technique is employed.

A short term moving average, usually three to six transactions, reflects the current movement of the measure and adjusts the current value by smoothing it with the last several transactions. A long term moving average, usually greater than ten transactions, indicates the overall trend of the account. These averages give a time dimension to each individual transaction.

A moving average analysis determined the number of transactions to apply to each measure. This resulted in selecting the moving averages presented in Table 1. Thus, the total number of measures used in the model development phase included the twenty-four sources and derived measures with their long term and short term moving averages. These sixty-one measures were computed for all transactions used in the analysis, both good accounts and fraudulent accounts, and were added to each transaction in the file that contained merchant information (e.g., type of vendor, location, etc.).

Measure	Moving Averages		Measure	Moving Averages		Measure	Moving Averages	
	Short	Long		Short	Long		Short	Long
1.	5	15	**9.**	4	15	**17.**	3	15
2.	6	20	**10.**	NA	NA	**18.**	6	20
3.	3	20	**11.**	3	15	**19.**	NA	NA
4.	3	15	**12.**	5	20	**20.**	NA	NA
5.	NA	NA	**13.**	3	15	**21.**	5	15
6.	3	15	**14.**	3	15	**22.**	NA	NA
7.	3	15	**15.**	NA	NA	**23.**	4	18
8.	3	20	**16.**	3	15	**24.**	7	25

NA=Not Applicable

Table 1: Moving Averages Selected for Each Measure

Model Development Process

The development of the model involved three sets of transactions. To simplify this description, a fraudulent transaction is referred to as a fraud transaction, and any transaction that is not identified as fraudulent is designated a good transaction. There is a possibility that a credit card transaction may be unobserved as fraudulent and thus be labeled as a good transaction. Research has addressed misclassification of training samples (Lacherbruch, 1966, 1974; Chhikara and Mckeon, 1984), but not in the context of fraud detection. Issues such as these were discussed in Chan and Stolfo (1998) and Provost and Fawcett (2001). The first group is composed of the fraud transactions that are randomly selected from the pool of all fraud transactions. Next, two good transactions were matched with each fraud transaction. The matching is done based on the amount and the type of purchase. The final group was composed of a random selection of good transactions.

Statistical Techniques

The process of developing the model to discriminate between normal account activity and fraudulent transactions involves five statistical procedures: (1) normality test, (2) T-test, (3) factor analysis, (4) stepwise discriminant analysis, and (5) discriminant analysis. Each of these techniques is discussed as it relates to the model development in identifying fraudulent activity.

A basic assumption is that the transaction amounts are normally distributed. The issue of normality was examined with plots of the purchase amount for sample accounts. The frequency charts illustrated that the measures do not have a normal shape and therefore are not normally distributed. The Shapiro-Wilks Normality Test (Sall, Creighton, and Lehman, 2005) used with less than 2,000 transactions was applied to formally determine if the measure was normally distributed. The results of this analysis clearly indicated that it was not a normal distribution. Using the 621,787 transactions from the good accounts, the KSL Normality Test (Sall, Creighton, and Lehman, 2005) is used with greater than 2,000 transactions and was applied to the transaction amounts with the same result.

Judicious use of transformations extends the range of applications in which normal sampling theory is appropriate. Transformations improve the normal approximation of many other distributions. The first step is to determine if an appropriate transformation exists to normalize the measure. After a review of potential distributions (Srivastaun, 2002) and the graphs of the actual data, the logarithm and the square root were applied to the transaction amount. The natural logarithm resulted in normalizing the transaction amount.

The T-test provides a statistical overview of the measures for both fraudulent and good transactions (Sall, Creighton, and Lehman, 2005). A by-product of this test procedure is the computation of the mean, standard deviation, standard error, minimum value, and maximum value. These statistics are used to validate the values for all measures and identify errors in the database. After correcting the errors, the T-test uses the mean and standard deviation for each group of transactions (fraudulent and good) to test the hypothesis that the population means are the same. The results of the T-test showed that the averages for all measures are significantly different between the good and fraudulent transactions. Thus, each measure has some degree of ability to discriminate between the good and fraudulent activity.

Factor analysis is an interdependence analysis used to determine the relationships among the different measures (McLachlan, 2005). The sixty-one measures created from the transaction data were derived to expand the original source information. In this analysis, the latent constructs underlying the measures are investigated. This yields insight into the different relationships that are really being monitored and identifies redundancies among measures.

The factor analysis resulted in twenty-four distinct groups, which reflect different aspects of purchase behavior. This suggests that a large number of measures are required to discriminate between good and fraudulent transactions. A sample of the logical groupings of the measures with their correlation values is given in Table 2.

Factor Analysis Group 1			**Factor Analysis Group 3**			**Factor Analysis Group 6**		
11	***SM***	***88***	12	LM	82	***11***	***SM***	***88***
4	SM	87	12	SM	80	18	LM	86
3	LM	85	***18***	***SM***	***76***			
1	SM	82	18	LM	75			
4	LM	79						
3	SM	75						

Note: SM Short Term Moving Average
LM Long Term Moving Average
Act Actual Value of Measure
Measure used in Final Model Indicated by ***Italics***

Figure 2: Partial List of Groups from the Factor Analysis

This information is useful in selecting a substitute measure that is easier to compute for the operational model. For example, the best measure to use in Group 1 is Measure 11's short term moving average. If Measure 11 is too difficult to calculate, Measure 4 that requires no calculations or prior transaction information is a potential substitute for it. As new measures were developed, this analysis also showed whether a new concept was unique or a variation of one in use.

Stepwise discriminant analysis techniques are designed to determine a good set of measures to use in identifying whether a transaction in fraudulent or not. The stepwise discriminant procedure (Jackson, 1991; Cooley and Lohnes, 1986) is a sequential process where the algorithm builds a model one measure at a time. Using forward recursion technique, the model initially has no measures and chooses the single most discriminating measure. On each succeeding step, the procedure selects the most discriminating measure from those remaining that add significantly to the model. The sequential nature of this process does not consider the effect of the simultaneous interaction from several measures. This provides the analyst with an extensive list of measures ranked in descending order from the most important to be investigated to the least. This information combined with the results from the factor analysis and T-test is utilized in developing the final discriminant models.

The discriminant analysis procedure requires that the analyst specify the exact set of measures to be used. The objective of the discriminant analysis process is to classify individual transactions using a set of independent measures, into one of two mutually exclusive and exhaustive categories. For example, on the basis of the Amount and High-Low Range over the last six transactions, the analyst wants to classify a transaction as either likely or unlikely to be fraudulent. This discussion is limited to two variables. It is expanded to more measures for the complete model.

The information presented in this example is hypothetical. Based on historical data, the analyst established the alert levels, the value above which the account's transaction is suspected of being fraudulent, at $325 for the purchase amount and $310 for the High Low Range. The 24th purchase was for $312 and increased the range to $298. Neither value considered individually exceeded the alert level and required no further action.

The discriminant analysis procedure displays this situation using both measures in a two dimensional view of the historical information. Figure 3 presents ten transactions that were classified as fraud, labeled with an **F**, and twenty-five **G**s identifying good transactions. Assuming that the individual alert levels are $325 on the purchase Amount and $310 on High-Low Range, the five fraud transactions in the shaded area appear on the analyst list for review. The other five fraud transactions in the white area were not identified for review, because they were below the alert value. Only three good transactions were incorrectly identified as fraudulent.

The discriminant procedure determines an equation or function, represented by a Classification Line, and separates the two sets; one dominated by fraud, the other by good transactions. In Figure 3, the line is positioned to separate the two groups with the fewest misclassifications as possible. Thus, if the line were moved down to correctly classify the **F** with the lowest value, then three good transactions would be incorrectly grouped. Similarly, three additional fraudulent ones would be missed if the line were moved upward to eliminate the misclassification of the two good transactions. Notice that two good transactions previously examined are now below the Line and ignored. With the Classification Line optimally placed based on historical information, the real power of this technique is that it provides the analyst with a single value that combines all the measures into one based on an equation that more accurately identifies fraud.

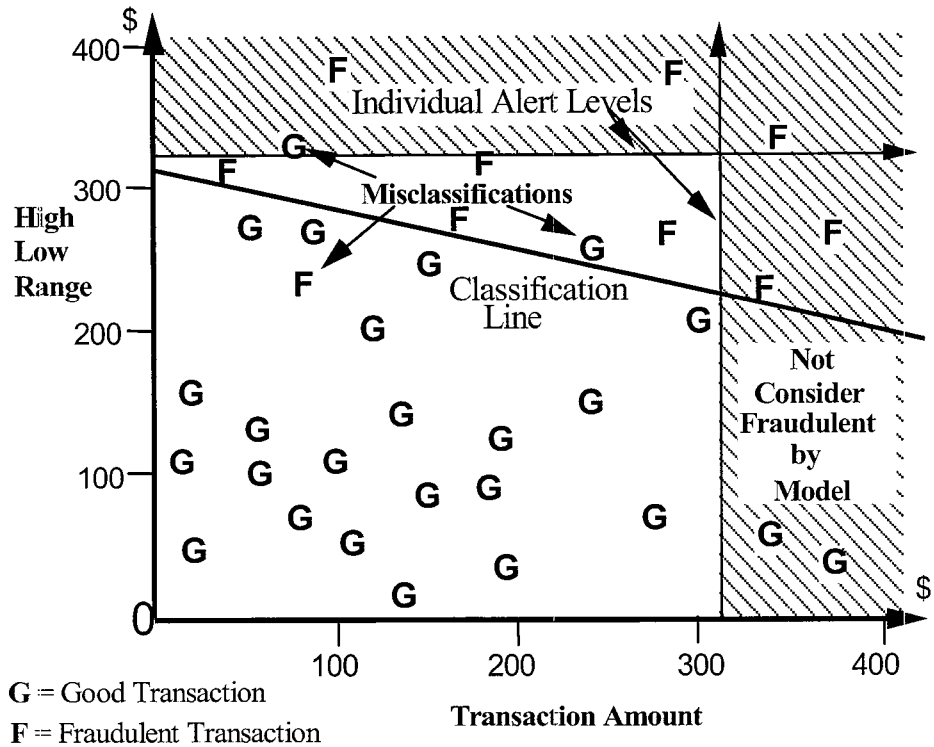

Figure 3: Discriminant Analysis Model Development

The discriminant function is then applied to the current transaction. For example, with the alert level for the individual measures at $325 and $310, the 24th purchase does not exceed either alert level and no action is taken (Figure 4). However, the Discriminant Function evaluates the interaction of the two measures and determines that the 24th transaction is above the Classification Line in the new Combination Alerts area. Therefore, it is categorized as suspicious. In addition, the probability of a fraudulent transaction assigns standard scale to the degree of suspicion. When the probability is greater than .500, it is called a fraudulent transaction.

The model computes the probability of fraud, classifies each transaction and identifies whether the discriminant function was correct. Figure 5 summarizes this information in the confusion matrix. The two distributions at the top of Figure 5 illustrate the two types of errors that are encountered in positioning the Classification Line: (1) good transactions that were classified as fraud, referred to as false positives, and (2) misclassified fraud transactions. The focus of the study is to capture as much of the fraud as possible, while not bothering too many cardholders who are using their cards legitimately. These are presented in the two shaded boxes.

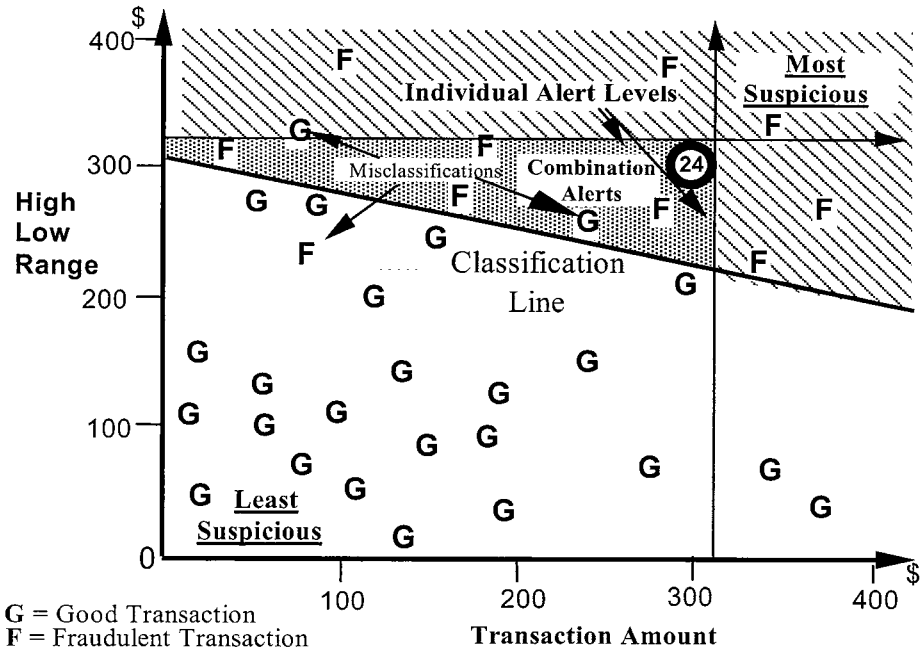

Figure 4: Discriminant Model with 24th Transaction

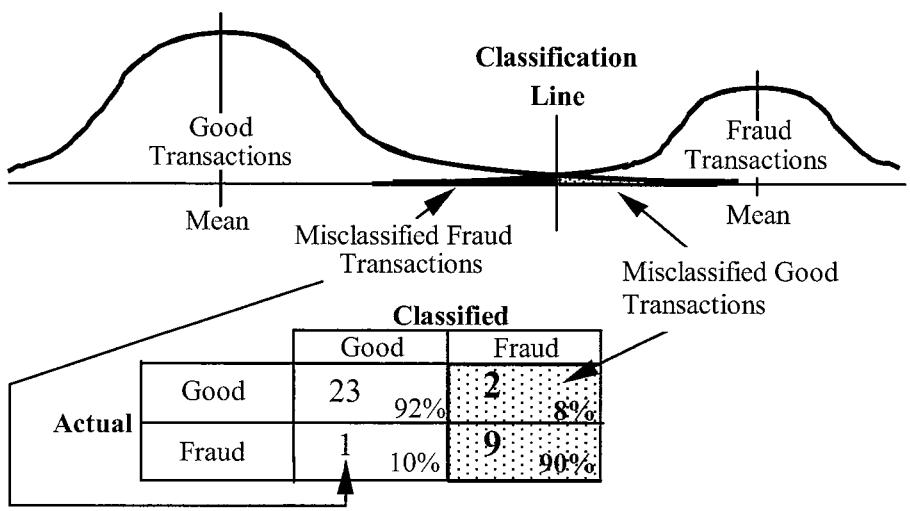

Figure 5: Confusion Matrix Development

Initially, all sixty-one variables were submitted with the complete analytic file and used in creating a baseline model. This model is an estimate of the performance that is possible with the final model. In reporting results for the segmentation analysis, the percentage of correctly classified fraud and false positives is given. The baseline model resulted in 4% false positives and 85% fraudulent transactions correctly identified. If the data is to be segmented, the subgroups must produce better results.

Two possible methods for segmenting required investigation. One method for separating the data was based on the dollar amount of the transaction, while the second used the sequence number. After examining the data, the logical values to segment using purchase amount were $26 and $120. When compared to the baseline, only a minor improvement is gained by segmentation on transactions by value.

The second method for separating the data is by the sequence number. With all of the short term moving averages fully calculated by the sixth transaction, this creates a logical separation point for segmentation. Additional points for segmentation are when the long term moving averages are calculated at the 15th and 20th transactions. All combinations of the segmentation using these points were examined. The best solution increases the effectiveness in identifying good transactions by 1% and fraudulent transactions by 6%. The data is segmented into three groups.

Low Number of Transactions	Transaction Sequence Numbers between 2 and 6
Medium Number of Transactions	Transaction Sequence Numbers between 7 and 19
High Number of Transactions	Transaction Sequence Numbers greater than 19

Ten of the original source and derived measures are required to achieve the best models for discriminating between transactions. The most important measure is Number 7. Measures 1, 3, 7 and 10 are related to velocity, the number of transactions in a period of time, while measures 4, 5, 6, 8, and 9 represent volatility, the change in transaction amounts. Concentration of purchases is given in measure 2. The final model combined two measures each of from velocity and volatility with the concentration variable.

RESULTS OF THE STUDY

The discriminate process constructs an equation that optimally divides the transactions in two groups and generates a probability that a given transaction is fraudulent. When the probability is greater than .500, it is classified a fraudulent transaction. However, the line that separates the two groups can be moved to a higher level than .50. Figure 6 illustrates the parallel Classification Lines generated for probabilities of .60 and .70. As the fraud probability is increased, the percentage of fraudulent transactions correctly identified and the number of good transactions falsely classified as fraudulent are reduced. The results of these different classification models are present in a confusion matrix. This is illustrated in Figure 7. The false positives are reduced from 8% as .50 to 4% at .60 and to 0% at .70, while a more dramatic decrease takes place in identifying the fraud as the probability increases.

The discriminant model development file provides an estimate of the effectiveness of each model. Figure 8 illustrates the impact from moving the fraudulent probability (Classification Level) higher. For each model, the "% Good" indicates the percentage of non-fraudulent transactions that are examined (false positives), and the "% Fraud" represents the percentage of correctly identified fraudulent transactions. For example, at .50 classification probability, 1.69% of good transactions are reviewed in order to capture 85.33% of the fraudulent ones. The number of false positives decreases as the classification value increases. Unfortunately, the number of fraudulent transactions also decreases. Thus, when the classification value increases to .55, the false positives decrease by .13% (1.69% - 1.56%), while the fraud captured decreases by 2.99% (85.33% - 82.34%). This tradeoff is given for each Classification Level in each of the three models. It is management's decision at what level to set the probability of classifying a transaction as fraudulent based on the tables.

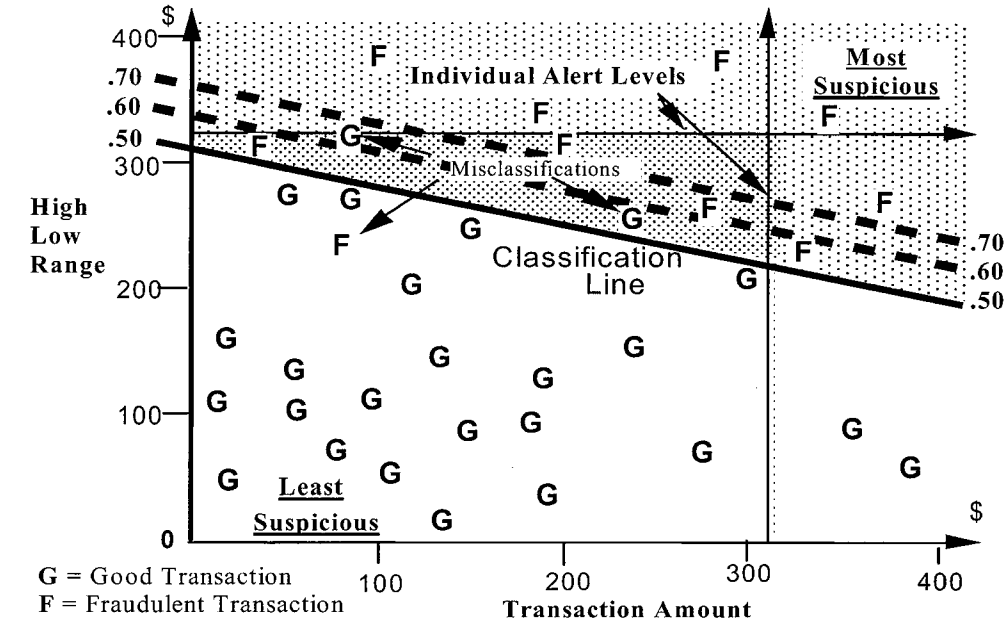

Figure 6: Different Classification Levels

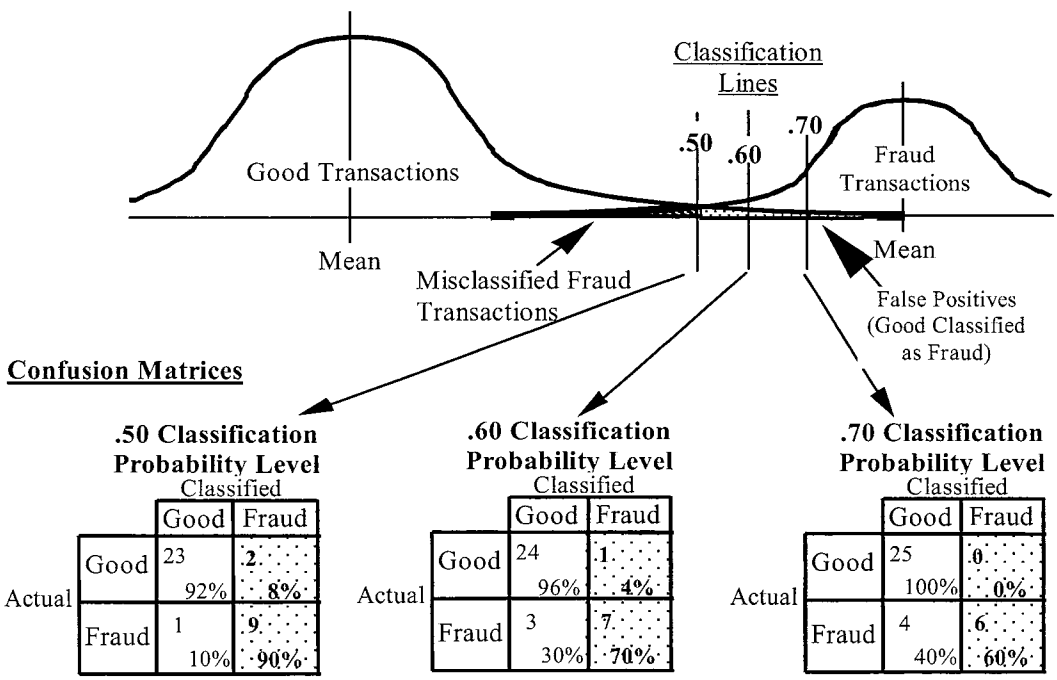

Figure 7: Different Classification Levels – Confusion Matrix

Classification	Low # of Transactions		Medium # of Transactions		High # of Transactions	
Level	% Good	% Fraud	% Good	% Fraud	% Good	% Fraud
0.50	1.69%	85.33%	7.79%	81.31%	3.02%	95.15%
0.55	1.56%	82.34%	6.66%	78.93%	2.41%	94.14%
0.60	1.36%	80.43%	5.52%	74.64%	2.04%	93.14%
0.65	1.23%	79.08%	4.30%	70.71%	1.58%	91.62%
0.70	1.10%	77.72%	3.40%	67.38%	1.11%	90.43%
0.75	0.91%	74.46%	2.31%	63.69%	0.79%	88.66%
0.80	0.65%	72.55%	1.89%	58.33%	0.42%	86.96%
0.85	0.32%	71.20%	1.37%	52.26%	0.28%	83.69%
0.90	0.19%	67.39%	0.85%	44.88%	0.28%	78.65%
0.95	0.13%	61.41%	0.52%	35.95%	0.09%	68.01%
Number of Transactions:						
Good	1539		2118		2154	
Fraud		368		840		1588

Figure 8: Results from the Discriminant Models

The discriminant function was set at a level where ten good transactions are misclassified for each fraudulent transaction identified. This results in capturing 32.3% of the fraudulent transactions and 30.4% of the fraudulent dollars.

The neural network model with two layers was developed using the five variables from the discriminant model with an additional 17 from the 61 original measures defined for the project and testing them with five million transactions. In tests (Figure 9) with another five million transactions, they were able to capture 28.9% of the fraudulent transactions and 27.8% of the dollars at a 10 to 1 ratio. Research continues to explain the difference in accounts identified between the two approaches.

	Number of Transactions	Value in Dollars
Discriminant Analysis	32.3%	30.4%
Neural Network	28.9%	27.8%
Combined	49.6%	46.1%

Figure 9: Final Summary

From the user's perspective, management prefers a model that is comprehensible. For many analysts, the discriminant model is easier to interpret than a neural network model. The measures that apply to a particular case may guide the subsequent investigation and prosecution. However, the immediate goal is to stop all fraudulent use of the card before the limit is exceeded. In order to provide a real time monitoring program, the models were simplified yielding a capture rate of approximately 42%.

CONCLUSION

The initial review of the transaction information in date order disclosed that 80% of the first purchases with the card were for gasoline at a safe, automatic pump. In many cases, several gas purchases were made within a few hours. It was assumed that the thief had friends fill their gas tanks using the card. This led to a velocity check rule to call the account owner if more than two gasoline transactions occurred within a specified number of hours.

A number of accounts produced a very high value on the suspicion rating during the testing phase, but when the owners were called, the cards were not stolen. These accounts were reviewed later to learn that the owners had filed for bankruptcy after exceeding the maximum limit on the card. The profile of theft and bankruptcy was the same for both types of fraudulent activity.

The number of calls to identify a fraudulent account was reduced to eight. The value of the call service is now justified since it cost $120 to reduce the loss of fraudulent transaction by $1,500 or more. The system is designed to stop the card if the suspicion score is high and allow the owner to call the company if the purchase it denied. A suitable compromise was established between stopping the card and calling account owner.

In this application, the speed of processing is of the essence. This is the case in credit card transaction processing where a high number of purchases are completed each day. Operationally, the discriminant model was modified to use measures that required the least amount of processing of prior transactions. The moving averages were changed to exponential smoothing equations to simplify the calculations. Similar modifications were made to the neural network model. These modifications are similar to those made by HNC Software as stated in their patent (Gopinathan, Biafore, Ferguson, Lazarus, Pathria and Jost, 1998). The company initially started the process to patent the models, but after the legal department described the amount of details that are required, the patent was not filed. There are over 80 patents that have been filed for fraud detection models (Provost, 2002). All models are constantly updated and modified to reflect the changing patterns of criminal behavior.

REFERENCES

Aleskerov, E., Freisleben, B. and Rao, B. (1997). CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection. Computational Intelleigence for Financial Engineering Proceedings of the IEEE/IAFE, IEEE, Piscataway, NJ, 220-226.

Barnett, V. and Lewis, T. (1994). Outliers in Statistical Data. New York: Wiley.

Bolton, R., and Hand, D.(2002). Statistical Fraud Detection: A Review, *Statistical Sciences 17*(3), 235-249.

Chan, P. and Stolfo, S. (1998). Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, 164-168.

Chhikara, R. and McKeon, J. (1984). Linear Discriminant Analysis with Misallocation of Training Samples. *Journal of. American Statistical Association 79*(388), 899-906.

Cooley, W. and Lohnes, P. (1986). Multivariate Data Analysis. Malabar, FL: Robert E. Krueger Publishing Company.

Ghosh, S. and Reilly, D. (1994). Credit Card Fraud Detection with a Neural Network. Proceedings of the 27th Hawaii International Conference on System Sciences, 621-630.

Gopinathan, K., Biafore, L., Ferguson, W., Lazarus, M., Pathria, A. and Jost, A. (1998). Fraud Detection Using Predictive Modeling. U. S. Patent 5819226, October 6, 1998.

Jackson, J. E. (1991). A User's Guide to Principal Components. New York: Wiley.

Lachenbruch, P. (1966). Discriminant Analysis When the Initial Samples Are Misclassified. *Technometrics 8*, 657-662.

Lachenbruch, P. (1974). Discriminant Analysis When the Initial Samples Are Misclassified II: Non-random Misclassification Models. *Technometrics 16*, 419-424.

Leonard, K.J. (1993). Detecting Credit Card Fraud Using Expert Systems. *Computers and Industrial Engineering 25*(1), 103-106.

McLachlan, G. (2005). Discriminant Analysis and Statistical Pattern Recognition. New York: Wiley.

Provost, F. (2002). Statistical Fraud Detection: Comment, *Statistical Sciences 17*(3), 249-251.

Provost, F. and Fawcett, T. (2001). Robust Classification for Imprecise Environments. *Machine Learning 42*, 203-210.

Sall J., Creighton, L., and Lehman, A. (2005). JMP Start Statistics, Third Edition. Toronto, Canada: SAS Institute/Thomson.

Srivastaun, M. (2002). Methods in Multivariate Statistics. New York: Wiley.

Yu, J., Qian, W., Lu, H., and Zhou, A. (2006). Finding Centric Local Outliers in Categorical/Numerical Spaces. *Knowledge and Information Systems 9*(3), 309.