Applied Intelligence 22, 47–60, 2005

A Data Mining Approach for Retailing Bank Customer Attrition Analysis

XIAOHUA HU
College of Information Science, Drexel University, Philadelphia, PA, USA 19104
xiaohua_hu@acm.org; thu@cis.drexel.edu

Abstract. Deregulation within the financial service industries and the widespread acceptance of new technologies is increasing competition in the finance marketplace. Central to the business strategy of every financial service company is the ability to retain existing customers and reach new prospective customers. Data mining is adopted to play an important role in these efforts. In this paper, we present a data mining approach for analyzing retailing bank customer attrition. We discuss the challenging issues such as highly skewed data, time series data unrolling, leaker field detection etc, and the procedure of a data mining project for the attrition analysis for retailing bank customers. We use lift as a proper measure for attrition analysis and compare the lift of data mining models of decision tree, boosted naïve Bayesian network, selective Bayesian network, neural network and the ensemble of classifiers of the above methods. Some interesting findings are reported. Our research work demonstrates the effectiveness and efficiency of data mining in attrition analysis for retailing bank.

Keywords: data mining, classification method, attrition analysis

1. Introduction

Deregulation within the financial service industries and the widespread acceptance of new technologies is increasing competition in the finance marketplace. Central to the business strategy of every financial service company is the ability to retain existing customer and reach new prospective customers. Data mining is adopted to play an important role in these efforts. Data mining is an *iterative* process that combines business knowledge, machine learning methods and tools and large amounts of accurate and relevant information to enable the discovery of non-intuitive insights hidden in the organization's corporate data. This information can refine existing processes, uncover trends and help formulate policies regarding the company's relation to its customers and employees [1]. In the financial area, data mining has been applied successfully in determining:

- Who are the likely attriters in the next two months?
- Who are likely to be your profitable customers?
- What is your profitable customers' economic behavior?
- What products are different segments likely to buy?
- What value propositions service different groups?
- What attributes characterize your different segments and how does each play in the person's profile?

In this paper, our focus is on applying data mining techniques to help retailing banks for the attrition analysis. The goal of attrition analysis is to identify a group of customers who have a high probability to attrite, and then the company can conduct marketing campaigns to change the behavior in the desired direction (change their behavior, reduce the attrition rate). In data mining based direct marketing campaign, it is well understood that targeting every customer is unprofitable and ineffective. With limited marketing budget and staff, data mining models are used to rank the customers and only certain percentage of customers are contacted via mail, phone etc. If the data mining model is good enough and target criteria are well defined, the company can contact a much small group of people with a high concentration of potential attriters [1–5]. The process of data mining for bank attrition analysis can be described in the following steps:

1. Problem definition: formulation of business problems in the area of customer retention.
2. Data review and initial selection
3. Problem formulation in terms of existing data
4. Data gathering, cataloging and formatting
5. Data Processing: (a) Data cleansing, data unfolding and time-sensitive variable definition, target variable definition, (b) Statistical analysis, (c) Sensitivity analysis, (d) Leaker detection, (e) Feature selection
6. Data modeling via classification models: Decision Trees, Neural Networks, Boosted Bayesian Networks, Selective Bayesian Network, an ensemble of classifiers
7. Result review and analysis: use the data mining model to predict the likely attriters among the current customers
8. Result Deployment: target the likely attriters (called rollout)

The paper represents a data mining approach for a retailing bank attrition analysis. The purpose is the identification of rules, trends, patterns and groups that can serve as potential indicators of attrition and identify the potential attriters in advance so the bank can take proactive actions to reduce the attrition rate. The paper is organized as follow: we first define the problem and formulation of business problems in the area of customer retention in Section 2. In Section 3, we discuss data selection, data review and initial, then data gathering, cataloging and formatting, data unfolding and time-sensitive variable definition. Then we discuss sensitivity analysis, leaker detection and feature selection. Next we describe data modeling via decision trees, neural networks, Bayesian networks, selective Bayesian network and an ensemble of classifiers with the above four methods in Section 4. In Section 5, we discuss the major findings, fields test results. Finally we conclude with our next step in Section 6.

2. Business Problem

2.1. *Brief Explanation of the Problem*

Our client is one of the top 10 retailing banks in the world. It offers many type of financial retail products to various customers. The product we discussed in this paper belongs to a certain type of loan service. Over 750,000 customers currently use this service with $1.5 billion outstanding, the product has had significant losses. Revenue is constantly challenged by a high attrition rate: every month, the call centers receive over 4500 calls from customers wishing to close their accounts. This, in addition to approximately 1,200 write-ins, "slow" attriters (no balance shown over 12 consecutive months) and pirated accounts constitutes a serious challenge to the profitability of the product, which totals about 5,700/month mostly due to rate, credit line, and fees. In addition to that, many customers will use the product as long as the introductory or "teaser" rate is in effect and lapse thereafter. There are customer management program costs and acquisition costs for each account: mailing costs $1/customer, telemarketing costs $5/customer. Cost of incentives (i.e. lowering rates to retain customer) needs to be considered and is dependent upon the offer. Our client didn't have a proactive or reactive retention effort. In most cases, it's come down to price reduction although there is a feeling that this need not be the only or most effective policy. However, the situation described above has motivated the business and technology executives of our client to review the possibility of setting a knowledge based retention effort through a combination of effective segmentation, customer profiling, data mining and credit scoring that can retain more customers, while maximizing revenue. The result of this initiative was the launching of the project described in this paper.

2.2. *Problem Definition*

The section presents the steps to understand and define the problem in terms of data available, time periods, and target fields. In this exercise, as in all data mining projects, the longest and most laborious part of the process is the selection, preparation and organization of the data [1, 6, 7].

There are different types of attriters in the product line:

- Slow attriters: Customers who slowly pay down their outstanding balance until they become inactive. Attrition here is understood comprehensively, where voluntary attrition can show more than one behavior.
- Fast attriters: Customers who quickly pay down their balance and either lapse it or close it via phone call or write in.
- Cross selling: Identify customers who are likely to purchase alternative products offered to existing loan customers such as life insurance and the like. The increase in relationships is believed to serve as a deterrent to attrition.

- High risk: Customers who are likely to become high risk.
- Pirating: Identify customers likely to transfer their relationship to competing products and away from our client.

These patterns are not unidimensional: a customer can display a subset of these behaviors over the life of the loan. At the same time, he/she can be influenced to change the behavior through the effective use of policies and incentives. Given this, the customer can be thought of as operating within a state diagram such as the one depicted in Fig. 1.

As the figure shows, a customer, through his actions, can migrate between activity and attrition where each state is defined in terms of a grouping of attributes. Based on this diagram, we decided to concentrate on two attrition problems, namely:

1. Utilizing data on accounts that remained continuously open in the last 4 months, predict, with 60 days advance notice, the likelihood that a particular customer will opt to voluntarily close his/her account either by phone or write-in.
2. Utilizing data on accounts that remained continuously open in the last 4 months, predict, with 60 days advance notice, the likelihood that a particular customer will have his account transferred to a competing institution. The account may or may not remain open.

The focus of the modeling process, and subsequent campaigns, will revolve around the resolution of the two classes of business problems related to improving customer retention and activation for the product line as identified by the business:

Problem Class #1: Retention of Existing Customers

The problem requires the stratification of customer segments by leveraging current segmentation model in order to:

- Develop models that predict the customers who are likely to attrite within 30 to 60 days on an ongoing basis.
- Identify the characteristics of the most profitable/desirable customer segments in order to develop policies to ensure their continued support, to grow the group, and to acquire more customers with similar characteristics.

Problem Class #2: Customer Activation Policies

Identify customer groups whose characteristics lend them to migrating from unprofitable/dormant to profitable. Once identified, the characteristics can enable the development of risk, maintenance and opportunity policies tailored to a successful migration.

3. Data Selection

Like all data mining exercises, the identification of relevant data in the right quantity and over a significant period of time is critical for the development of meaningful models [1, 8, 9]. Given this and worked with the domain experts, we proceeded to identify the necessary data sources available and those readily accessible for initial review. Table 1 summarizes the data sources identified and their expected functions.

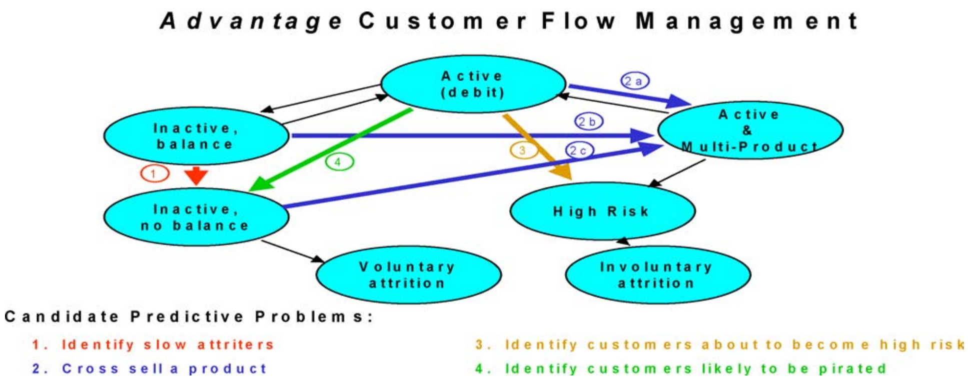

Figure 1. Customer's attrition state diagram.

3.1. Data Preprocessing Goals

The data preprocessing state consists of the series of activities necessary to create a compacted file that:

- Reflects data changes over time.
- Recognizes and removes statistically insignificant fields
- Defines and introduces the "target" field
- Allows for second stage preprocessing and statistical analysis.

This was accomplished through three steps, detailed in the sections below:

- Time series "unrolling"
- Target value definition
- First stage statistical analysis

3.1.1. Data Period Identification. Given the data availability and its time periodicity, we decided to start the data selection by extracting a subset of records of accounts whose status was "Open" i.e. currently appearing as bona fide users for the period of **12/2001–3/2002**. In addition, and for comparison purposes, we also requested data on accounts whose status is either "Closed" or "Charged Off" on a data not prior to 11/2001. This will allow us to review the characteristics of attriters and enable the creation of a voluntary attriter. Based on this consideration, 45,814 random records were selected, each representing one account of which 42,547 are open and the remainder, or 3267 have been closed since Dec, 2001. The records obtained are stored in a temporary table to be used to join with all relevant tables that will produce, for every account, the status of the account FOR EACH MONTH for the past 4 months. This means that we are able to reproduce ALL THE FIELDS for each month stating in December 2001 and not only those for March, 2002.

3.1.2. Time Series "Unrolling". The data extracted from the data source had one row per account for each month. In our application, historical customers records are used to group customers into two classes—those who are attriters and those who are not. In order to save space, every month a query checks every field against the previous month. If there is no change, no rows are added and the value of Effective Start Date (EFF_START_DT) remains as that during which a change was last recorded (which is the same as "a new row was inserted"). If any attribute changes, a whole new row is added with the corresponding EFF_START_DT updated. Therefore, it is very likely that some of the accounts will have less than the corresponding number of months in cases where no activity is recorded. For example, if an account has had no activity since December '2001, the last row will be the one for that month and it is up to the user to extrapolate it all the way to the current month. Understanding this when arranging the files is critical to developing the attrition model.

Table 1. Description of identified (Potential) data sources for related data.

Data source	Description & relevance to attrition modeling
DDS warehouse	Credit Card Data Warehouse containing about 200 product specific fields. Originating at various points (Write-ins, external database, scores etc.), the data is compacted according to a set of operational rules that reduce size for non-changing fields. The Warehouse contains 6 months of data and is rotated on a monthly basis. In some cases, additional attributes allow for data to cover up to 18 months. For the current exercise, the period includes 4 month history information. The Credit Card Data warehouse is the primary source of data for retention modeling problems.
Third party data	A set of account related demographic and credit bureau information. The data is available from an external provider, Dynamark, Inc. The data is linked to DDS Warehouse data to provide additional predictive data
Segmentation files	Set of account related segmentation values based on our client's segmentation scheme, which combines Risk, Profitability and External potential. The segment data combine with DDS data extract to overlay with results of models.
Payment database	Database that stores all checks processed. The database can categorize source of checks. This data set allows the finding of date, balance and originator of checks and is used to identify pirated accounts

Table 2. Naming convention for time sensitive DDS data for the 4 months period.

Period	Nomenclature
Last month (March 2002)	T0_CURRENT_BALANCE
One month back (Feb 2002)	T1_CURRENT_BALANCE
Two month back (Jan 2002)	T2_CURRENT_BALANCE
Three month back (Dec 2001)	T3_CURRENT_BALANCE

Since data records in the data file is one row per account for each month, we need to use the last 4-month historical information to build a predictive model. To reflect the data changes over time and maintain the seasonal behavior of the data, it is necessary to combine the 4 individual monthly data files into one data file, in which each account has one row with all the last 4 months financial information. The data format used is required for the implicit data to be made explicit and the time periods to be itemized into individual fields. To accomplish this, we classify the variable into static variables and time sensitive variables [1]. The values of static variables do not change over time. Examples of this are: Account Number, Mother's Maiden name, Address, and the like. The values of time sensitive variables changes from month to month and it is essential to retain all these different values for the last 4 months in order to find some seasonal/temporal behavior related to the account activity. The time sensitive variables were assigned a time prefix (T0 means the last month, T1 one month back from the last month, T2 two months back from last month and T3 three months back from the last month). So, for example, the variable *CURRENT_BALANCE* for the period of December 2001 to March 2002 is redefined in Table 2.

Given this, the next task consisted of generating the additional fields on the "clean" formatted files and adding them to the resulting file. After the time-series data unrolling (Table 3), the fields in the data set increased from 250 to around 870.

3.1.3. Target Value Definition. Like many real data mining applications, normally there is no data mining target field defined directly in the data warehouse. It is part of the data mining procedure to define the proper target field based on the business objective for the data mining analysis. With the help of the business domain experts, we define the target value in terms of existing data and, with these, define the value of the target variable, i.e., the variable that determines the voluntary attriters, hereby defined as **VA_ACCTS (1: attriter, 0: non-attriter)**. It is defined in terms of:

1. Status code (*CRD_ST_CD*)
2. Status change date (*CRD_STATUS_CHANGE_DATE*)
3. Closed reason code (*CRD_CLS_REA_CD*)

The formula for definition is:

CRD_ST_CD = C (Closed) &&
CRD_STATUS_CHANGE_DATE Between *Dec-1-2001* and March-31-2002 &&
CRD_CLS_REA_CD (reason code) in
[0 1 23 25 26 28 29 30 35 36 40 41 42 80 81 82 83 84 97 98 31 32 33 34]

The reason codes for a voluntary attriter (customer requested) are: "0 1 23 25 26 28 29 30 35 36 40 41 42", the reason codes for a voluntary attriter (customer requested) related to pricing issues are: "31 32 33 34". According to this definition, the average attrition rate for the section of the data received is 2.2% of all customers for a 4-month period.

3.1.4. First Stage Statistical Analysis. Of the 870 fields in the time-series data set, a significant portion

Table 3. Data file after time-series unrolling.

Account#	Name	…	T0_Current_balance	T1_Current_balance	T2_Current_balance	T3_Current_balance	…	…
1	Jim	…	1234.7	2345.6	1489.87	234.9	…	…
2	Tom	…	0.0	0.0	769.09	0	…	…
…	…	…	…	…	…	…	…	…
15000	Ana	…	200.0	215.0	250.0	278.0	…	…
…		…	…	…	…	…	…	…
45814	Mike	…	780.0	1200.90	300	100.0	…	…

Table 4. Data file after target field defined.

Account#	Name	...	T0_Current_Balance	T1_Current_Balance	T2_Current_balance	T3_Current_Balance	...	VA_ACCTS
1	Jim	...	1234.7	2345.6	1489.87	234.9	...	0
2	Tom	...	0.0	0.0	769.09	0	...	1
...	...	...	...	...	...	...	...	...
15000	Ana	...	200.0	215.0	250.0	278.0	...	0
...		...	...	...	...	...	...	...
45814	Mike	...	780.0	1200.90	300	100.0	...	0

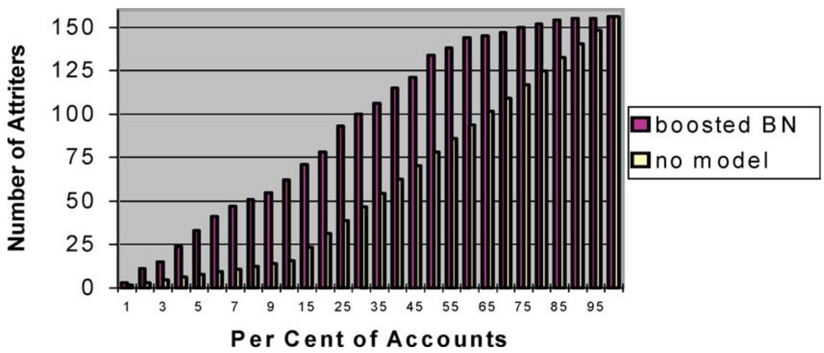

Figure 2. Boosted Naïve Bayesian model lift chart.

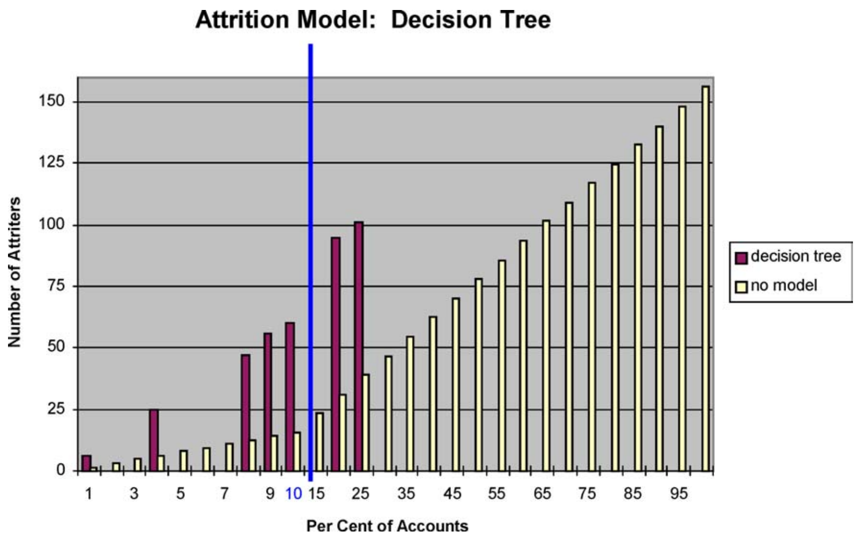

Figure 3. Decision tree model lift chart.

of them is constants, null fields. Filtering them out in the early stage can dramatically reduce the data mining time and improve the model accuracy. The statistical analysis, the first in a series, is done in order to obtain an initial understanding of the data quality: number of unknown fields, relative frequency, early indicators, averages and target data distribution. As an initial field discrimination step, the fields where a single value appeared in more than 99.95% of all records or null was deemed statistically

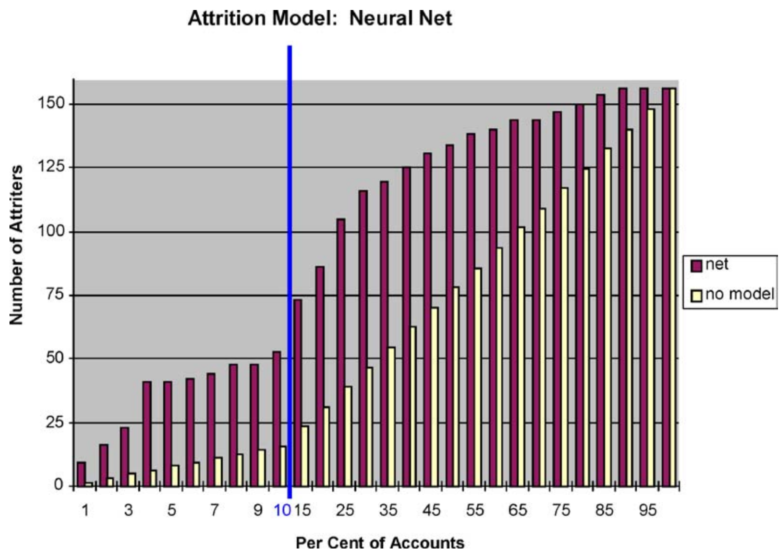

Figure 4. Neural net model lift chart.

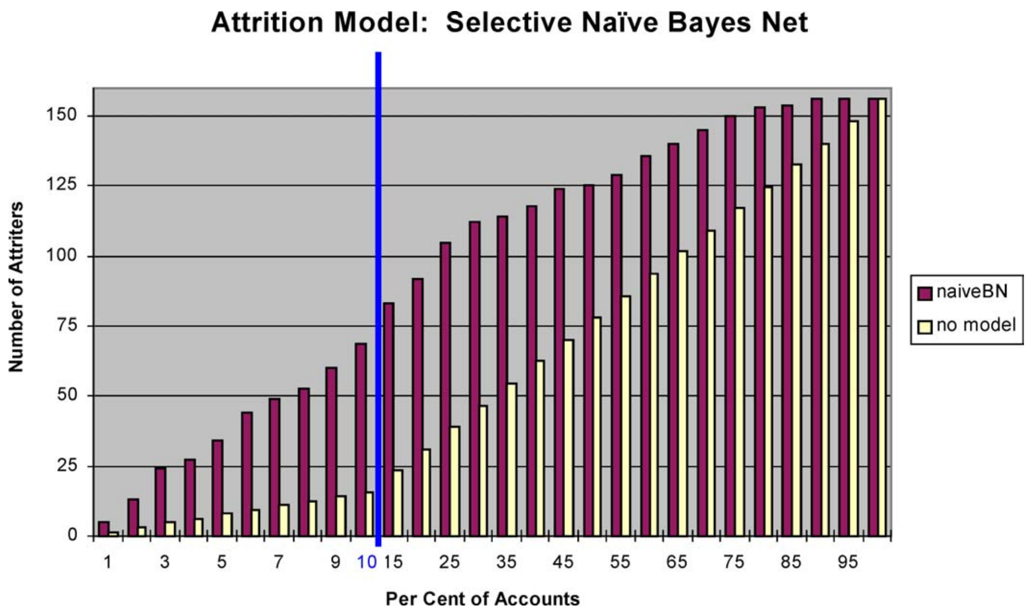

Figure 5. Selective Naïve Bayesian network model lift chart.

insignificant and removed from the set of attributes. These fields are removed from both the data and meta-data files to ensure their removal from the modeling process, thus reducing the computing time required. This process reduced the fields from 870 to 655.

3.2. *Data Premodeling*

The data premodeling stage is the next critical step in the generation of the files used for modeling. This stage consists of three main steps, namely: (1) field sensitivity analysis to filter fields with low correlation to target the field and detect data "*leakers*", (2) field reduction to create a compacted file with highly relevant fields, (3) file set generation of all balanced and unbalanced sets required for training, testing and iterative verification of results and model refinement.

3.2.1. Field Sensitivity Analysis and Field Reduction. The field sensitivity analysis is used to determine each attribute's "contribution" to the modeling process. We use a rough set based field selection algorithm [4]. Our algorithm generates a merit value for each field, consider the interaction/correlation among the fields and identify a minimum subset of fields from the entire field. On the other hand, if a field whose merit value is too high, it is considered to be a potential *leaker* [10]. Leakers are fields that "leak" information on the target. For example, a field with a value representing account closure could leak information on attrition, and would confound modeling efforts. While some leakers are readily explained, many times they are included in business rules whose relation to the target is not apparent. In this case, the best way to determine if a field is indeed a leaker is to discuss the findings with those familiar with the data schema and the business problem. Some of the leaker fields we identified from the data sets are *bankruptcy score, risk level, bankruptcy reason code, number of time the person used the cards, number of sale, the closed reason code* etc. In many circumstances, field names and values are not always representative of their function, and need clarification. One the other hand, fields that are suspected but turn out *not* to be leakers constitute potential predictors in the model. In some cases, the values for a field are constant (i.e., have a standard deviation of zero) and thus the merit value is zero. These fields should be removed in order to improve data mining processing speed and to generate better models. After the leaker field analysis, we applied our novel feature selection algorithm [4] on the data set to identify a minimum subset of features related to the target field. For example, through this effort, the initial set of 655 attributes in the data set was reduced to 242 after processing.

3.2.2. Files Set Generation. Our sample file comprises of 45814 records, the average monthly attrition rate is 0.55%, and the accumulated attrition rate of the last 4 months is around 2.2%. In order to build a good model from this highly skewed data set, we need to build a more balanced representation of attriters and non-attriters in the training data set. The reason is that in the original data file, we have high non-attriters percentage (97.8%) vs. a very low attriter rate (2.2%); a learning model can achieve high accuracy by always predicting every customer to be a non-attriters. Obviously, such a high accurate model is useless for attrition analysis [10, 11]. We created a random sample file where we include about 938 attriters and then we add enough non-attriters into it to make it a dataset with 50–50 percentage of each class category (attriters vs. non-attriters), then file was divided into *balanced, train* and *test* files as well as *raw* (i.e., unbalanced) *test* and *held aside* files for verification purpose. The *balanced train file* consisted of 50% of the records containing target values, i.e., for whom VA_ACCTs = 1. The *balanced test, raw test*, and *raw held aside files* consisted of approximately 1/6 of the targets each. As defined earlier in Section 3.1.3, targets in the raw files represent 2.2% of the total number of records for the files being reviewed. These files were handed over to the data mining component for further statistical analysis, data mining and clustering work.

4. Data Mining Model Development Process

4.1. Evaluation Criterion: Lift

As pointed in [4, 5, 11, 12], prediction accuracy, which was used to evaluate the mining algorithms, is not a suitable evaluation criterion for the data mining applications such as attrition analysis. The main reasons are:

1. Classification errors (false negative, false positive) are treated equally, but in attrition analysis, false positive and false negative have different impact and different consequence, they should be dealt with differently,
2. Accuracy is used to measure the performance of the learning algorithm on the whole data set, The goal of the attrition analysis is not to predict the behavior of every customer, but find a good subset of customers where the percentage of attriter is high. For

attrition analysis, the data set is highly skewed and noisy, it is very difficult to build a model with good accuracy.

In attrition analysis, our goal is to use history information to build an effective data mining model and then use the data mining model to predict the most likely attriters and then take proactive action to prevent the customer attrition. So it is required that learning algorithms need to classify with a confidence measurement, such as a probability estimation factor or certainty factor (also called scores in attrition analysis). The scores will allow us to rank customers for promotion or targeting marketing. Lift instead of the predictive accuracy is used as an evaluation criterion. As pointed in [9], if the data mining model is good enough, we should find a high concentration of attriters at the top of the list and this higher proportion of attriters can be measured in terms of "lift" to see how much better than random the model-based targeting is. Generally, lift can be calculated by looking at the cumulative targets captured up to $p\%$ as a percentage of all targets and dividing by $p\%$ [9]. For example, the top 10% of the sorted list may contain 35% of likely attriters, then the model has a lift of $35/10 = 3.5$. A lift reflects the redistribution of responders in the testing set after the testing examples are ranked. After the learning algorithm ranks all testing examples from most likely responders to least likely responders, we divide the ranked list into some deciles (the top 10% is finer partitioned in our test experiments: we measure the lift in each percentage), and see how the original responders distributed in these deciles. Lift measures the increased accuracy for a target subset based on a model-scored ranked list. Using past information collected over several months on usage of the financial service, our task is to build a model for predicting the customer class in the next two months and apply it to the whole customers. The prediction model is used to rank the customers based on their likelihood of attrition. As shown in section, the attrition rate for our clients is low (2.2%) and it is difficult or impossible to predict with high accuracy for all customers, and usually it is not necessary to predict all the customers because in practice, for attrition analysis, it is a good practice to contact a small percentage of customers and hope this small percentage of customers contains a high concentrated percentage of attriters than random sample.

4.2. Data Mining Models Based on Different Algorithms

We are interested in models that maximize lift. A good model in our analysis should concentrate the likely attriters near the top in the sorted list based on the attrition scores generated by the model. We need to use learning algorithms that can produce scores in order to rank the testing examples. Algorithms such as Naïve Bayesian, decision tree, neural network satisfy our requirement. We performed several data mining analysis using four different data mining algorithms and an ensemble of classifiers of the above 4 algorithms [4]. These are:

1. Boosted Naïve Bayesian (BNB)
2. Neural Network (NeuralWare Predict: a commercial neural network from NeuralWare Inc)
3. Decision Tree (C4.5 with some modifications for scoring and ranking [10])
4. Selective Naïve Bayesian (SNB).
5. An ensemble of classifier of the above four methods [4]

4.2.1. Boosted Naïve Bayesian Networks (BNB). The BNB data mining method combines boosting and naive Bayesian learning [6, 13–15]. Boosting is a general method of improving the predictive accuracy of any two-class learning algorithm, which works in successive stages. In the first stage, all the training examples are weighted equally and the two-class learning algorithm is used to acquire a classifier. In the second stage, the examples that are misclassified by this first classifier are upweighted, and a second classifier is learned that focuses on these examples. In the third stage, the examples misclassified by the second classifier are upweighted, and a third classifier is learned. The boosting process can be repeated for as many stages as desired. Applied with naive Bayesian learning, generally five to twenty stages are beneficial. The results described here use just five stages.

Like other software, our BNB software identifies which attributes are most predictive of an example being a target. Unlike most other software, BNB reports which values (or numerical ranges) of an attribute are most predictive. For example, BNB automatically identifies that the value 2 of the attribute T1_CRD_ACCOUNT_FORMAT is an important predictor. According to the supplied documentation, this value 2 signifies "account which has been active but is currently not active." Also unlike other software, BNB evaluates the statistical significance of the predictors that it reports. The significance of a predictor depends on both its lift (i.e. predictive benefit) and of its coverage (i.e. number of examples to which it applies). BNB does not report predictors that may be spurious, because they have low coverage or low lift. The lift chart of BNB is shown in Fig. 2.

Results

Pct	Cases	Hits Boosted BN	% hits	Lift	Hits no model
1	70	3	4.3	1.9	1.5
2	141	11	7.8	3.5	3.1
3	212	15	7.1	3.2	4.7
4	283	24	8.5	3.9	6.2
5	354	33	9.3	4.2	7.8
6	425	41	9.6	4.4	9.3
7	496	47	9.5	4.3	10.9
8	567	51	9.0	4.1	12.5
9	638	55	8.6	3.9	14.0
10	709	62	8.7	4.0	15.6
15	1063	71	6.7	3.0	23.4
20	1418	78	5.5	2.5	31.2
25	1772	93	5.2	2.4	39.0
30	2127	100	4.7	2.1	46.8
35	2481	106	4.3	1.9	54.6
40	2836	115	4.1	1.8	62.4
45	3190	121	3.8	1.7	70.2
50	3545	134	3.8	1.7	78.0
55	3900	138	3.5	1.6	85.8
60	4254	144	3.4	1.5	93.6
65	4609	145	3.1	1.4	101.4
70	4963	147	3.0	1.3	109.2
75	5318	150	2.8	1.3	117.0
80	5672	152	2.7	1.2	124.8
85	6027	154	2.6	1.2	132.6
90	6381	155	2.4	1.1	140.4
95	6736	155	2.3	1.0	148.2
100	7091	156	2.2	1.0	156.0

Variables of Interest

There are 14 most significant positive predictors of the target class picked up by BNB. The top 4 attributes are listed as below in order. Each predictor is a certain value (or numerical range) of a certain attribute. A value of "z" means zero in the original dataset. "counts" is the number of targets versus non-targets with this value of the attribute. "zscore" is the measure of statistical significance.

- Attribute 84 T0_CURRENT_BALANCE {Current Balance carried in hundreds of cents} between −1840.52 and 1277.62: counts 209/86, odds 2.43418, zscore 7.17529
- Attribute 119 T1_CRD_ACCOUNT_FORMAT {Record Format of the Account. Values are: 1 = Never-Active Account, 2 = Account which has been active but is not currently active, 3 = Currently Active Account, 4 = Delinquent Account} between 1.9 and 2.2 : counts 281/154, odds 1.82764, zscore 6.10613
- Attribute 56 T0_NON_CF_LS_MIN_PY_DUE value z {This figure corresponds to the minimum payment due on the last statement. it is used in conjunction with accrued arrears and the number of cycles delinquent to permit automatic delinquency adjustment.} counts 353/214, odds 1.65221, zscore 5.8568
- Attribute 40 T0_NON_CF_LS_OS_BAL {The actual ending balance as it appeared on the cardholder's last statement. this field is not affected by adjustments.} between −1840.52 and 1277.62: counts 189/98, odds 1.93171, zscore 5.38532

4.2.2. Decision Trees. Decision tree methods build a collection of rules for use as a predictive model [10, 16, 17]. The advantage of this approach is that the rules are easy to understand, and they are frequently useful for discovering underlying business processes. The disadvantage of decision tree approaches is that these models usually do not perform as well as other models. We have developed a proprietary modification for standard decision tree algorithms for use in "lift" problems [10] where, for example, we want to minimize performance in the top 25% of the predicted data (and care less about performance elsewhere). This is the situation for common problems, such as attrition and targeted mailings. The lift chart of decision tree is shown in Fig. 3.

Result

PCT	Cases	Hits decision tree	% hits	Lift	Hits no model
1	70	6	8.6	3.9	1.5
4	283	25	8.8	4.0	6.2
8	567	47	8.3	3.8	12.5
9	638	56	8.8	4.0	14.0
10	709	60	8.5	3.8	15.6
20	1418	95	6.7	3.0	31.2
25	1772	101	5.7	2.6	39.0

Some of the rules are:

Rule 8: (Lift = 5.347, 1-Cover = 0.029)
T0_CF_HD_ACT_MNTHS ≤ 2
T3_CRD_SGMNT_CD = A1
-> class 1 [0.889]
Rule 12: (Lift = 4.102, 1-Cover = 0.162)
T0_CF_CURRENT_BALANCE ≤ 407.06
T2_CF_DATE_LAST_STATEMENT ≤ 1998.055
T3_CRD_SGMNT_CD = A2
-> class 1 [0.859]
Rule 14: (Lift = 3.927, 1-Cover = 0.318)
T2_CRD_ANN_CHRG_DT ≤ 1998.164
T0_CF_YTD_NET_NO_PURCHASE ≤ 0
T0_CF_CURRENT_BALANCE ≤ 407.06
T3_CRD_SGMNT_CD = A1
-> class 1 [0.812]
Rule 9: (Lift = 3.868, 1-Cover = 0.385)
T0_CF_CURRENT_BALANCE ≤ 407.06
T3_CRD_CR_BUR_SCR > 606
T3_CRD_SGMNT_CD = A3
T3_CRD_BKRPCY_REA_CD_3 > 9260
-> class 1 [0.889]

4.2.3. Neural Networks. Neural networks are a well-established approach for modeling data [7, 14, 18]. The advantage of this approach is that neural network models tend to be among the most predictive models. The disadvantage of neural network models is that it can be harder to understand their output. For our work we have used a commercial package (NeuralWare Predict) that:

- Selects appropriate input transfer functions for fields
- Selects subsets of the variables to model the data, and
- Builds "constructive" neural network models.

The lift chart of neural network is shown in Fig. 4.

Results

PCT	Cases	Hits Neural Net	% hits	Lift	Hits no model
1	70	9	12.9	5.8	1.5
2	141	16	11.3	5.2	3.1
3	212	23	10.8	4.9	4.7
4	283	41	14.5	6.6	6.2
5	354	41	11.6	5.3	7.8
6	425	42	9.9	4.5	9.3
7	496	44	8.9	4.0	10.9
8	567	48	8.5	3.8	12.5
9	638	48	7.5	3.4	14.0
10	709	53	7.5	3.4	15.6
15	1063	73	6.9	3.1	23.4
20	1418	86	6.1	2.8	31.2
25	1772	105	5.9	2.7	39.0
30	2127	116	5.5	2.5	46.8
35	2481	120	4.8	2.2	54.6
40	2836	125	4.4	2.0	62.4
45	3190	131	4.1	1.9	70.2
50	3545	134	3.8	1.7	78.0
55	3900	138	3.5	1.6	85.8
60	4254	140	3.3	1.5	93.6
65	4609	144	3.1	1.4	101.4
70	4963	144	2.9	1.3	109.2
75	5318	147	2.8	1.3	117.0
80	5672	150	2.6	1.2	124.8
85	6027	154	2.6	1.2	132.6
90	6381	156	2.4	1.1	140.4
95	6736	156	2.3	1.1	148.2
100	7091	156	2.2	1.0	156.0

4.2.4. Selective Naïve Bayesian Networks. The naive Bayesian classifier is a probabilistic, predictive model that assumes that all attributes are conditionally independent of each other given the target variable [19] i.e. within each class, the attributes are unrelated. The naïve Bayesian classifier is simple, inherently robust with respect to noise, and scales well to domains that involve many irrelevant features. Moreover, despite its simplicity and the strong assumption that attributes are independent within each class, it has been shown to give remarkably high accuracies in many natural domains. The selective naive Bayesian classifier that we used is an extension to the naive Bayesian classifier designed to perform better in domains with highly correlated (redundant) features. The intuition is that, if highly correlated features are not selected, the classifier should perform better given its feature independence assumptions. Attributes are selected by starting with an empty set of attributes, and then incrementally adding that single attribute (from the set of unselected attributes) the attribute that most improves the accuracy of the resultant classifier on the test set. Attributes are selected until the addition of any other attribute results in a fall in accuracy of the classifier. The lift chart of the selective Bayesian networks is shown in Fig. 5.

Results

PCT	Cases	Hits SelectiveBN	% hits	Lift	Hits no model
1	70	5	7.1	3.2	1.5
2	141	13	9.2	4.2	3.1
3	212	24	11.3	5.1	4.7
4	283	27	9.5	4.3	6.2
5	354	34	9.6	4.4	7.8
6	425	44	10.4	4.7	9.3
7	496	49	9.9	4.5	10.9
8	567	53	9.3	4.2	12.5
9	638	60	9.4	4.3	14.0
10	709	69	9.7	4.4	15.6
15	1063	83	7.8	3.5	23.4
20	1418	92	6.5	2.9	31.2
25	1772	105	5.9	2.7	39.0
30	2127	112	5.3	2.4	46.8
35	2481	114	4.6	2.1	54.6
40	2836	118	4.2	1.9	62.4
45	3190	124	3.9	1.8	70.2
50	3545	125	3.5	1.6	78.0
55	3900	129	3.3	1.5	85.8
60	4254	136	3.2	1.5	93.6
65	4609	140	3.0	1.4	101.4
70	4963	145	2.9	1.3	109.2
75	5318	150	2.8	1.3	117.0
80	5672	153	2.7	1.2	124.8
85	6027	154	2.6	1.2	132.6
90	6381	156	2.4	1.1	140.4
95	6736	156	2.3	1.1	148.2
100	7091	156	2.2	1.0	156.0

4.2.5. A Hybrid Approach: An Ensemble of Classifiers. An ensemble of classifiers is to generate a set of classifiers instead of one classifier for the classification of new object, hoping that the combination of answers of multiple classifiers result in better accuracy [4, 15, 20, 21]. Ensemble of classifiers has been proved to be a very effective way to improve classification accuracy because uncorrelated errors made by the individual classifier can be removed by voting. A classifier, which utilizes a single minimal set of classification rules to classify future examples, may lead to mistakes. An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way to classify new examples. Many methods for constructing ensembles of classifiers have been developed, some are general and some are specific to particular algorithms [4, 8]. We adopted a hybrid approach: we first built 4 classifiers using Boosted Naïve Bayesian (BNB), NeuralWare predict, Decision Tree, Selective Naïve Bayesian (SNB), then we ensembled a classifier from the 4 classified based on the majority vote of them [4]. The ensembled classifier does have a better lift measure than any of the 4 individual classifiers.

Results

Pct	Cases	Hits ensemble of classifier	% hits	Lift	Hits no model
1	70	7	10.0	4.7	1.5
2	141	16	11.3	5.2	3.1
3	212	23	10.8	4.9	4.7
4	283	42	14.8	6.8	6.2
5	354	42	11.9	6.8	7.8
6	425	43	10.1	5.4	9.3
7	496	48	9.7	4.4	10.9
8	567	52	9.3	4.2	12.5
9	638	61	9.6	4.4	14.0
10	709	63	8.9	4.0	15.6
15	1063	81	7.7	3.5	23.4
20	1418	96	6.5	3.0	31.2
25	1772	104	5.9	2.6	39.0
30	2127	121	5.7	2.6	46.8
35	2481	131	5.3	2.4	54.6
40	2836	144	5.1	2.3	62.4
45	3190	153	4.8	2.1	70.2
50	3545	154	4.4	2.0	78.0
55	3900	155	4.0	1.8	85.8
60	4254	156	3.6	1.7	93.6
65	4609	156	3.3	1.5	101.4
70	4963	156	3.1	1.4	109.2
75	5318	156	2.9	1.3	117.0
80	5672	156	2.7	1.2	124.8
85	6027	156	2.5	1.1	132.6
90	6381	156	2.4	1.1	140.4
95	6736	156	2.3	1.0	148.2
100	7091	156	2.2	1.0	156.0

5. Data Mining Findings

The initial studies unveiled a number of relationships between variables as well as threshold values that justify further discussion and analysis. Following is a summary of the more salient points and their possible meaning:

Variables	Results & implication
Most recent Current Balance *(CURRENT_BALANCE)*	The most recent current balance showed a strong predictive value when the amount fell below approximately $1000.00. A small but significant segment was of those with negative balances, i.e., of customers who overpay. Review of accounts whose previous balance falls below the threshold may be candidates for proactive action. Such candidates can also be "Possibilities" subjects during inbound calls. It can also be an indicator of "Balance Attriters" in the case of negative balances.
Current Balance with constant values *(CURRENT_BALANCE=12 NON_CF_CURRENT_BALANCE=15)*	The accounts with values of $12.00 and $15.00 dlls. Showed prominently among the results for attrition. If these are interest charges or related to annual charges, it could hint at policies for retention/exiting/ win back of customers.
Segment *(CRD_SGMNT_CD)*	The association of a group to a specific segment (as defined in the DDS Data Warehouse) was a significant value for segments A1–A4. With a larger sample group, we intends to allocate groups to segments (as defined by the Marketing group) in order to run more focused models.
Annual charge date *(CRD_ANN_CHRG_DT)*	The billing period within the first trimester of the year is predictive of impending attrition for customers with reduced balance *(CURRENT_BALANCE* < 407). The results point an attrition pattern for lagging users of "zero balance" users who take the charge as a disincentive to maintain the product.
Number of payments *(NO_PY)*	Accounts with a number of payments made over the same billing period or payments made to cover low balances over a continuous period *(YTD_NET_PURCHASE_AMT* ≤ 62.0) can be predictive of attrition. A request for Payoff could indicate, for some accounts, a likelihood of closing (due to a recent annual charge or competitor's bid) which could be averted.
Incentive Interest *(INCENT_PRI_ANN_MRCH_R)*	Incentive pricing appeared to be somewhat predictive at the value of 4.9%. This result may warrant a more in-depth study of segment-based review of "rate chaser" population.

The table above shows that several specific values (or ranges of numerical values) of several attributes are useful predictors of retention and/or attrition. These explanations increase our confidence that these values of these attributes will continue to be predictors in the future.

Field Test

To test the effectiveness of the data mining models, our client conducted a field test on their customers. The test wanted to show two points:

1. The top percentage of the customer attrition list does contain concentrated attriters,
2. The data mining based marketing approach is effective for retention purpose.

They ran the model generated from the ensemble of classifiers approach on the current customers and then sorted the customers based on the attrition scores. They decided to contact the top 4% of the existing customers (around 750000) from the list, which has around 30000 customers. They divided the customers into 2 groups randomly, each with 15000 customers and took different proactive actions to each group: for group1, the marketing department contacted each customer and offered some incentive packages to encourage the customers to stay with the company, for group 2, there is no contact. After two months later, they examined the list and found out, for group 1, the attrition rate is very low (0.12%), for group two, the attrition rate is very high, almost 5.6%, while the accumulated average attrition rate for two month is 1.1%, thus achieved a lift of 5.0 (consistent with the list of 4.6 in the test data set). The lower attrition rate among group 1 did indicate, if the proactive action is in time and proper, it does have an impact on the customers' behavior, the high attrition rate among group 2 demonstrate that our data mining model is accurate and the top 4% captured a high concentrated proportion of attriters.

6. Conclusion

In this paper, we present a data mining approach for retailing bank customer attrition analysis. We discuss the challenging issues such as highly skewed data, time series data unrolling, leaker field detection etc, and procedure of a data mining task for the attrition analysis for retailing bank. We discuss the use of lift as a proper measure for attrition analysis and compare the lift of data mining model of decision tree, boosted naïve Bayesian network, selective Bayesian network, neural network and the ensemble of class of the above methods. Our initial findings show some interesting results. The field test conducted by our clients proved that the data mining prediction model for attrition is very accurate and the target-oriented campaign is very effective. Based on above results and new source files available on segmentation; we plan to review the voluntary attrition trends on a segment-by-segment basis. A thorough clustering study is also planned for the data to review the natural grouping of the data and how it lines up with the segmentation in terms of incidence, variables and number of groups.

References

1. A.M. Hughes, *The Complete Database Marketer: Second-generation Strategies and Techniques for Tapping the Power of Your Customer Database*, McGraw Hill, 1996.
2. B. Baesens, M. Egmont-Petersen, R. Castelo, and J. Vanthienen, "Learning Bayesian network classifiers for credit scoring using Markov chain monte Carlo search," Tech report, Utrecht University, Institute of Information and Computer Science, Netherland.
3. S. Bhattacharya, "Direct marketing response models using genetic algorithms," in *Proc. of the 4th International Conference on Knowledge Discovery and Data Mining*, 1998, pp. 144–148.
4. X. Hu, "Using rough sets theory and database operations to construct a good ensemble of classifiers for data mining application," in *Proc. of the 2001 IEEE International Conf. on Data Mining*, 2001, pp. 233–240.
5. F. Provost and T. Fawcett, "Analysis and visualization of classifiers performance: Comparison under imprecise class and cost distribution," in *Proc. of the 3rd International Conference on Knowledge Discovery and Data Mining*, pp. 43–48.
6. M.Berry and G. Linoff, *Mastering Data Mining: The Art and Science of Customer Relationship Management*, John Wiley & Son, 1998.
7. T. Masters, *Neural, Novel & Hybrid Algorithms for Time Series Predictions*, John Wiley & Sons Inc, 1995.
8. N. Indurkhyn and M. Sholom, "Solving regression problems with rule-based ensemble classifiers," in *Proc. of the ACM SIGKDD*, 2001, pp. 287–292.
9. G. Piatetsky-Shapiro and B. Masand, "Estimating campaign benefits and modeling lift," in *Proc. of the 5th SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 185–193.
10. X. Hu, "Comparison of classification methods for customer attrition analysis" in *Proc. of the Third International Conference on Rough Sets and Current Trends in Computing*, Springer, 2002, pp. 487–492.
11. C. Ling and C.Li, "Data mining for direct marketing: Problem and solutions," in *Proc. of the 4th International Conference on Knowledge Discovery & Data Mining*, 1998, pp. 73–79.
12. B. Masand and G. Piatetsky-Shapiro, "A comparison of approaches for maximizing business payoff of prediction models," in *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 195–201.
13. C. Elkan, "*Boosted* and naïve bayesian learning," Technical Report No. CS97–557, University of California, San Diego, Sept. 1997.
14. M. Klenin, "A simple neural network predictor for volatile time series data nonlinear financial forecasting," in *Proc. of the 4th International Conference on Neural Networks in the Capital Market*, 1996, pp. 134–140.
15. R.E. Schapire, Y. Freund, P. Bartlett, and W.S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," *The Annals of Statistics*, to appear.
16. J.R. Quinlan, "Induction of decision tree," *Machine Learning*, vol. 1, no. 1, pp. 81–96.
17. J.R. Quinlan, "Bagging, boosting, and C4.5." in *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, 1996, pp. 725–730.
18. G. Grudnitski and A. Quang, "Important factors in neural networks' forecasts of gold futures prices," in *Proc. of the 4th International Conference on Neural Networks in the Capital Market*, 1996, pp. 163–176.
19. N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, pp. 131–163, 1997.
20. R.E. Schapire, "The boosting approach to machine learning: An overview," MSRI Workshop on Nonlinear Estimation and Classification, 2002.
21. R.E. Schapire, M. Rochery, M. Rahim, and N. Gupta, "Incorporating prior knowledge into boosting," in *Proc. of the Nineteenth International Conference on Machine Learning*, 2002.
22. X. Hu and N. Cercone, "Data mining via discretization, generalization and rough set feature selection," *Knowledge and Information System: An International Journal*, vol. 1, pp. 33–60, 1999.