Management Dynamics

Volume 15 | Number 2 Article 4

December 2015

Leveraging Business Analytics for Marketing Decisions in Retail Banking

Prakash Singh
Accounting and Finance, Indian Institute of Management, Lucknow, India

LEVERAGING BUSINESS ANALYTICS FOR MARKETING DECISIONS IN RETAIL BANKING

Dr. Prakash Singh*

ABSTRACT

In a service based economy, companies strive to continue deriving revenue by creating and nurturing long term relationship with clients. A case in point is Retail Banking, where customer value is of utmost importance. In today's hyper-competitive environment, banks are aggressively leveraging their customer base to engage in revenue driving activities such as cross selling and up- selling. To be successful, it is imperative for banks to embrace the power of analytics to gain insights and appropriately evaluate risks and opportunities- enabling more efficient decision making in the quest to enhance share of the wallet. This paper examines the various applications of analytics in Retail Banking and provides pointers for analytic implementations. The paper is divided into 10 parts. Part 1 is Introduction, Part 2 is about Business Analytics, Part 3 is about Data Mining, Part 4 deals with Literature Review, Part 5 is about Theoretical Framework on Analytics in Retail Banking, Part 6 is Methodology and Discussion on results, Part 7 is Inference and Conclusion.

Key Words: *Data Mining, Cluster Analysis, Retail Banking, Credit Scoring*

INTRODUCTION

The Indian banking sector is rapidly growing and globalizing, making it imperative for Indian banks to ensure that their products, processes and practices match with those of the best banks in the world. One of the major drivers of the Indian banking industry in the recent past is the emergence of the Retail banking. This phenomenal growth in Retail banking in India is mainly attributable to fast growth of personal wealth, disposable income, favorable demographic profile, rapid development in information technology, financial market reforms, and several micro-level supply side factors.

But despite this rapid surge in retail banking, it is still plagued by its inability to stand out in an increasingly competitive and commoditized marketplace (Haenlein et al., 2007). The banking industry started to face a set of new challenges that has an overall negative impact on industry's margin and profitability. The major key challenges which banks currently started facing were- a highly saturated market where products and prices were no longer the key differentiators, thus pushing up retention costs (World Retail Banking Report, 2013); lack of personalized connectivity with customers due to

---

****Associate Professor, Accounting and Finance, IIM, Lucknow. He can be contacted on his email address p_singh@iiml.ac.in***

diminishing role of branch banking at an alarming rate; new delivery channels facing inconsistency; thereby, resulting in disjointed experience for many customers (World Retail Banking Report, 2013).

Thus, in the wake of these challenges it became crucial for banking organizations to focus on effective ways to build customer trust and drive stable long term growth through improved customer experience and tailored offerings. According to IBM's 2010 Global Chief Executive Officer Study, 89 percent of banking and financial markets CEOs believed that top priority of banks is to understand, predict and give customers what they want as building and sustaining profitability over long term requires an ability to increase wallet share, improve customer satisfaction and loyalty.

Therefore, serving mass market customers more cost effectively and anticipating customer needs assumes more significance (IBM Software, Business Analytic Report, 2011). Today the customers are more empowered and have more choices in terms of financial service providers. They are savvier, price sensitive and far less loyal. They know how they want to be treated and expect their banks to know this as well. Banks that do not or cannot understand customer needs face a variety of challenges that directly impact their profitability.

Hence, to stimulate increased levels of convenience to customers and to offer superior services and to reduce the number of negative experiences, banks are exploring new strategies and technologies like Business Analytic (BA, henceforth) to establish a competitive advantage over others, to facilitate improved decision making and to optimize business processes (Watson and Wixom 2007).

BUSINESS ANALYTICS

In recent years, the advent of Business Analytic tool adoption has transformed the way marketing is done and how banks manage information about their customers. There has been a strong interest in the use of BA systems to provide benefits to the organization (Davenport and Harris 2007; Davenport et al. 2010) as it constitutes an important component for organizational success (Watson, 2013). BA signifies applying various advanced analytic techniques to data to answer questions or solve problems. It is not a technology in and of itself, but rather a group of tools that are used in combination with one another to gain information, analyze that information and predict outcomes of the problem solutions (Bose, 2009). Another simpler view of BA was presented by Watson, (2013) by comparing it with Business Intelligence (BI)and highlighted the difference between the two; as "BI is getting data in (to the warehouse) and getting data out (data access and analysis) whereas analytics is the analysis part of BI. Thus, analytics is the algorithms (example, neural networks) and methods used to find patterns in data (example, customer segmentation analysis) or to optimize performance (example, revenue management).

BA constitutes of four stages – Decision Support System, Data Warehouse Management, Data Mining and Knowledge Management.

1. ***Decision Support System:*** Making important decisions through the use of information technology. It assists and supports managers in making decisions and keep them connected to the decision making loop. It improves decision making process and increases the efficiency of decision making.
2. ***Data Warehouse Management:*** Can be conceptualized as a process of centralized data management and retrieval. It is the core of a well-developed BI program(Bose, 2009).
3. ***Data Mining:*** Can be conceptualized as the automated extraction of hidden predictive information from databases. In other words, it is the process of analyzing large data sets in order to find

patterns that can help to isolate key variables to build predictive models for management decision making (Bose, 2009).

4. ***Knowledge Management:*** It allows informed decision making in which organization's best practices for each decision making process are pushed to the desktops of end-users as embedded logic within analytic applications. These applications are typically powered either by business rules engines (which apply logical conditions to determine how a certain case should be handled) or predictive models (which probabilistically identifies the most likely action to achieve the desired results) (Bose, 2009).

Thus, BA influences decision making through collection, storage and interpretation of large amounts of high quality data stored in a data warehouse by using descriptive, predictive and prescriptive analytics.

Descriptive Analytics: Objective is to describe 'what has occurred'. The tools used for this analytics are reporting, Online Analytical Processing (OLAP), dashboards/scorecards, and data visualization (Watson, 2013).

Predictive Analytics: Objective is to focus on 'what will occur' in the future. The algorithm and methods used are regression analysis, machine learning and neural networks.

Prescriptive Analytics: Objective is to show 'what should occur'. It is used to optimize system performance. Through a combination of forecasting and mathematical programming, the prices are dynamically set for the good over time to optimize revenues(Watson, 2013).

Many banks across the globe have started focusing on the application of BA, but despite this increased attention, it is still difficult to get a holistic view of how BA can and should be used in organizations (Watson, 2013). Based on this issue we have addressed in this paper a holistic view of how BA adoption can leverage benefits in the context of marketing decisions to the retail banking industry.

DATA MINING CONCEPT

The term data mining means – "to extract useful information from large datasets" (Hand et al., 2001). In other words it is, "the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques." Gartner Group (2004). It consists of building a model from data (Carrier and Povel 2003) which can perform one or more of the following types of data modeling: Association, Classification, Clustering, Forecasting, Regression, Sequence Discovery and Visualization. But according to Wang and Wang (2008), most of the value of data mining comes when it is used for predictive modeling. It generates predictive models automatically, which can help banks to predict how much profit prospects and customers will provide and how much risk will entail from fraud, bankruptcy, charge-off and related problems (Bose, 2009) but the choice of data mining technique is based on the data characteristics and business requirements (Carrier and Povel 2003). Data mining is mainly used to extract the 'gold hidden' in a company's data, but, it addresses only a very limited part of a company's total data sets. According to Weiss et al, (2005) about 90 percent of a company's data are never being tapped or looked at like- letters from customers, emails, correspondence, recording of phone calls with customers, contracts, technical documentation, patents and so on. But, through the use of advanced data mining tool which consists of – text mining and web mining (Hearst, 2003; Fan et al. 2006) banks can dig out the hidden 'gold' from these unstructured information sources.

The generation of predictive models through data mining also known as predictive analytics is a dimension of business analytics that allows organizations to assess both risks and opportunities. From a retail banking perspective it provides answers to questions such as "which customers are likely to default on loans? Which are likely to be profitable, long-term customers?" Getting the right answers to these questions is important because it has a direct effect on the banks success. To answer such questions a historic data is used to construct a model that correlates the characteristics of a group of customers with their financial behavior. From a large set of measurements, key influencing factors are identified. The same information is then collected about other customers or prospects and matched against a profile that correlates with the target behavior. Banks can make decisions about issuing loans or marketing new products based on expected consumer behavior (Lamont, 2005).

3.1 Data Mining Tasks

Data mining tasks are used to extract patterns from large datasets. According to Fayyad et al., (1996) pattern extraction is an important component of any data mining activity and it deals with relationships between subsets of data. The identification of patterns in a large data set is the first step to gaining useful marketing insights and making critical marketing decisions (Shaw et al., 2001). The data mining tasks can be broadly divided into five categories, namely: Dependency Analysis, Class Identification, Concept Description, Deviation Detection and Data Visualization (Shaw et al., 2001) and which specific task to be used is determined by the marketing problem in hand.

3.1.1Dependency Analysis:

Dependency knowledge is the association between sets of items stated with some minimum specified confidence (Agarwal et al. 1993) also known as "market basket analysis" (Berry et al., 1997) details out the relationship between different products purchased by a customer. It is used for developing marketing strategies for promoting products.

3.1.2 Class Identification:

It groups customers into classes, which are defined in advance. There are two types of class identification tasks- Mathematical taxonomy and concept clustering.

Mathematical Taxonomy- Algorithms produce classes that maximize similarity within classes but minimize similarity between classes (Frawley et al., 1992). For example in retail banking a bank can classify its customers based on their income or past purchase amounts and then design its marketing strategies and target customer accordingly.

Concept Clustering- It determines clusters according to attribute similarity as well as conceptual cohesiveness as defined by domain knowledge. Users provide the domain knowledge by identifying useful clustering characteristics (M.J. Shaw et al., 2001). For example, based on the session log data of internet banking, bank can classify its customers according to its internet banking activity like, frequent transactions through internet, seldom and no transaction through internet.

3.1.3 Concept Description:

Concept description is a technique to group customers based on domain knowledge and the database, without forced definitions of the groups. Concept description can be used for summarization, discrimination, or comparison of marketing and customer knowledge. Data summarization is the process of deriving a characteristic summary of a data subset that is interesting with respect to domain knowledge and the full data file. Using summarization, a marketer can learn about customer characteristics by grouping them according to their occupation, income, spending patterns and types of purchases, and build customer profiles. Discrimination describes qualities sufficient to differentiate

records of one class from another. Comparison describes the class in a way that facilitates comparison and analysis with other records (Frawley et al., 1992, Shaw et al., 2001)

3.1.4 Deviation Detection:

Deviations are useful for the discovery of anomaly and changes. Anomalies are things that are different from the normal. It can be detected by analysis of the means, standard deviations, and volatility measures from the data. In addition to anomalies, variables or attributes may have significantly different values from the previous transactions for the same customer or group of customers. For example, a bank may find a sudden increase in the credit purchases of an individual customer. This change can be due to change in customer's status or change in income and not necessarily a fraud (Shaw et al., 2001).

3.1.5 Data Visualization:

Data visualization software allows marketers to view complex patterns in their customer data as visual objects complete in three dimensions and colors. They also provide advanced manipulation capabilities to slice, rotate or zoom the objects to provide varying levels of details of the patterns observed. To explore the knowledge in database, data visualization can be used alone or in association with other tasks such as dependency analysis, class identification, and concept description and deviation detection (Shaw et al., 2001).

3.2 The Key Application Areas of Data Mining (Predictive Analytics):

The application of analytics has increasingly become a key differentiator and contributor of competitive advantage in banks. The banks around the globe are applying analytics for focusing and consolidating data from various entities to gain customer insights for cross-selling and to better serve customer needs. The three broad categories of analytics application in retail banks around the world are- Financial Analytics; Risk Analytics; and Customer Analytics.

Financial Analytics- The purpose of financial analytics is to discover how the banks have performed relative to its competitors, and the levers to push for greater output. For example, a bank would want to determine the customer segments which provide the bank with the most profits, how the bank's customer base is spread geographically, and which product brings in the most profit. The insights gained are used to invest in creating capabilities which will enable the bank to differentiate itself in particular geographies (Asian Banker Research, 2013).

Risk Analytics- It addresses business risks by analyzing risks associated with customer credit, regulations and financial risk factors through modeling. The main purpose is to improve credit performance and reduce the level of non-performing loans (NPLs) by examining customer creditworthiness. It creates predictive models that can be used to determine the probability of fraud and propensity to default (Asian Banker Research, 2013).

Customer Analytics- Is the application of analytics to mine customer data to discover patterns and correlations through customer segmentation. The insights generated are used for building deeper customer relationships; offering best values to the customers and optimizing the management of financial transactions as well as purchase and investment decisions (Asian Banker Research, 2013).

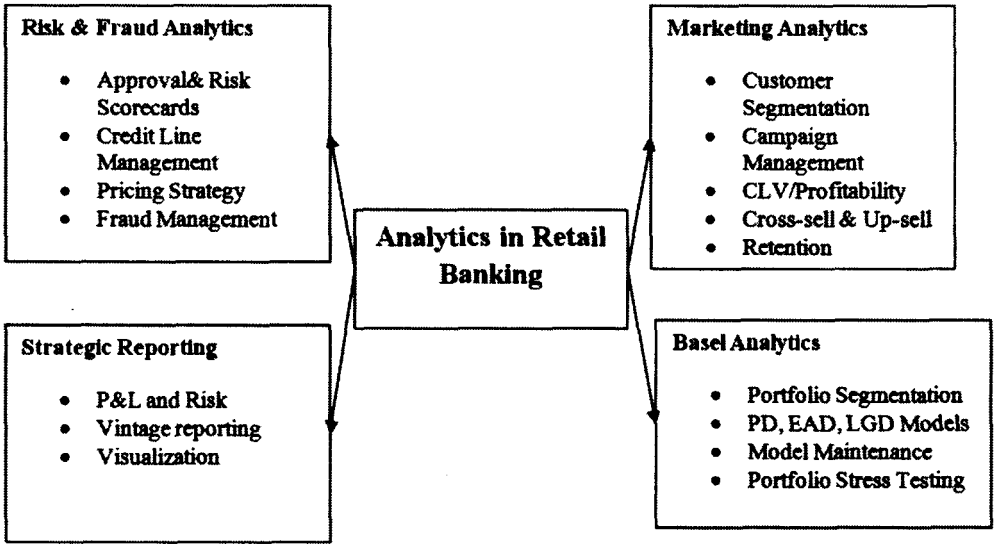

Figure1-Key Application Areas of Data Mining

LITERATURE REVIEW

The extant literature on Business Analytics shows that what differentiates companies in today's highly competitive markets is their ability to make accurate, timely, and effective decisions at all levels- operational, tactical and strategic- and to address their customers preferences and priorities (Bose, 2009). Almost all the industries around the globe have started using advanced analytics to analyze their data (both structured and unstructured), and by combining information on past circumstances, present events, and projected future actions (Apte et al., 2003). A growing number of evidence shows that analytics generates considerable business value as according to IBM Institute for Business Value and MIT Sloan Management Review, 2010 report, which surveyed nearly 3000 executives and business analysts about their companies use of analytics, reveals that top performing companies use analytics five times more than low performing ones (LaVallel et al., 2010).

Another study in 2011 showed that more than 4,500 executives, managers, and business analysts discovered that the number of companies using analytics to create a competitive advantage had surged by 57 percent in the past year and the performance gap had widened between those companies that rely on advanced analytics and those don't (Kiron et al., 2011). Furthermore, Brynjolfsson, Hitt and Kim (2011) studied 19 large publically traded firms and founded that firms that emphasizes decision making based on data and analytics have output and productivity that are 5-6 percent higher than what would have been expected without this emphasis. They also found that the relationship between the use of data and analytics appeared in other performance measures also such as asset utilization, return on equity, and market value.

However, there are some organizations that have not been successful in utilizing BA to increase their profit and achieve their expected performance (Gesner et al. 2005). Such mixed results have motivated researchers to examine a variety of factors that contribute to the successful implementation of BA (Jordan et al. 2008; Jukic, 2006; Fan et al. 2006).

Indeed, one of the significant themes that have emerged in the research for successful BA adoption is that it must suit the problem space, or decision environment within which it is used (Clark et al. 2007). However, past academic researches and publications indicate that this success has yet not been realized in many organizations and that the user has not been able to make a connection between its BA capabilities and the decision environment (Hostmann et al. 2007). Till now the research has focused on the theoretical and computational process of pattern discovery and a narrow set of applications such as fraud detection or risk prediction. Therefore, against this backdrop, we revisited extant perspectives on BA adoption in firms, by specifically focusing on the retail banking perspective. We also acknowledge in our study the important role played by marketing decisions in the current customer-centric environment and recognize the need for an integrated framework for efficient flow of customer knowledge through the application of BA tools for building effective marketing strategies.

Therefore, the backdrop of our study specifically focuses on the following issues:

1) A holistic view of how BA adoption can leverage benefits in the context of marketing decisions to the retail banking

2) How Data Mining (predictive analytics) tool is used and helps for determining new product development, promotion, recommending products and cross-sell and up-sell.

3) Identifying the possible returns from BA adoption in retail banks in the context of marketing decisions.

The conceptual analysis of the extant perspectives of BA presented in this paper makes an important contribution to the literature on BA by demonstrating that the efficacy of data mining tasks and techniques and knowledge management can influence bank's performance positively by integrating the knowledge discovery process with the management and use of the knowledge for marketing strategies (Shaw et al., 2001).

DATA MINING MODELS FOR RETAIL BANKING INDUSTRY

In this section, let us see in general what all analytic techniques are available and how they might be helpful for the retail banking industry. Data Mining Algorithms are generally classified into following 2 categories:

- Supervised Learning
- Unsupervised Learning

Supervised learning is directed towards predicting a previously known target variable. The objective is to determine the target variable as a function of a set of independent variables. On the other hand unsupervised learning is non-directed, and there is no previously known result. It is descriptive in nature. Here, we will discuss most common supervised and unsupervised data mining techniques used in retail banking industry.

6.1 Decision Trees

These are used for supervised learning wherein information is derived through comprehendible rules in the form if-then-else structures. Decision tree using CART (Classification and Regression Tree) algorithm is commonly used for customer retention, fraud prevention, etc. The probability of a customer churning and that of fraud are set as target variables in above-mentioned applications respectively.

A decision tree is a tree like graph, which has variable cut offs at each branching. They classify instances in 2 categories, positive and negative. They can become very large especially if there are a lot of input variables, and hence are required to be pruned such that accuracy and complexity are traded off.

1.1. Regression

This is pretty much the regression we study in quantitative and statistical sciences. The applications are pricing and costing of products such that maximum profit is attained considering the fact the pricing will impact no. of customers. The purpose is to identify the appropriate relationship among the variables, and then trade off high price vs. no. of customers to achieve highest profitability. This is a supervised learning method.

1.2. Logistic Regression

This supervised model is again used for predicting a categorical dependent variable based on other independent predictor variables. The regression curve is not a straight line, but a logistic curve given by following formula:

$$P = e^{f(x)}/(1+e^{f(x)}) Or \quad Log_e f(x) = P/(1-P)$$

This method is commonly used for credit approval and fraud prevention. Decision trees and logistic regression can be used interchangeably but effectiveness of both varies with scenarios and business problems at hand.

1.1. Clustering Model

The simplest definition of clustering would be grouping similar data together. This is an unsupervised learning technique. The main logic is that the difference between 2 clusters should be maximum and within a cluster minimum. This difference is quantified using distance metric or similarity metric. Customer segmentation is the most common use of clustering; however it can also be used for fraud detection.

Other than the above mentioned models, Support Vector Machines, Probability Density Estimation, Apriori association, etc. are some of the data mining models, which can be utilized in retail banking industry.

CREDIT SCORING AND DEFAULT PREDICTION MODELING

Credit scoring helps lenders take decisions on whether to lend or not, depending on borrowers credit rating. It is much more difficult to score in a retail set up where banks just have information as provided by the customer and transactional history. Individual credit scores are not provided by agencies as in case of corporate customers. Default prediction can also help banks to proactively identify default accounts and take appropriate actions so as to reduce bad debts. Such prediction also helps in identifying how to deal with such accounts, what parameters are causing those issues and what can be done to prevent the customer from defaulting.

Here, a sample prediction modeling approach is represented on a commonly available German loan default data.

7.1 Preliminary Research

For any model to succeed, it is important to ponder upon what kind of information we have. A number of banks (viz. Axis Bank, SBI, ICICI Bank, HDFC Bank) and their loan application forms were studied so as to find out what kind of information a bank has about its borrower. This information is listed below:

- **Personal Information:** Ids, category, marital status, number of dependents, residence ownership and duration, permanent address, etc.
- **Educational Information:** Past and future plans, qualification
- **Employment Information:** Employment nature and length, designation, total employment period, Organization nature, name and address.
- **Financial Information:** 3 Year income tax returns, balance sheets, income statement, and bank statement, Monthly income, bank account details, investment details, active loan information, credit card details
- **Property/Lifestyle Related Information:** Vehicle owned and hypothecated to, electronics owned, etc., property details,
- **Co-applicant/Guarantor Information**

While some information might not provide any insight and is kept only for record purpose like id, category, etc., others can provide as a starting point for pattern discovery in terms of both behaviour (credit card transactional details, and lifestyle products owned) and general attribute (income, age, dependents, etc.). This data along with the web transaction history, ATM data, bank transaction history, and social media data can provide means to uncover defaulters.

7.2 About the Data

The German credit score data is a common data used for modeling loan default predictions. It contains data of 1000 loan accounts, with 20 variables. In the model, 70% data is taken to train the model, and 30% to validate it.

7.3 Methodology

Following steps are incorporated to model the default prediction for the data at hand:

7.3.1 Business Objective and Corresponding Data Mining Objectives

The business objective for a credit-scoring model can vary across departments in their purpose and approach. A few examples are as discussed below:

- Identification of credit worthy customers for targeted marketing
- Decision regarding whether to lend or not, ad how much to lend if yes
- Predict future behaviour regarding default of a customer
- To reduce defaults from current portfolio of loans through proactive assistance to customers

All above business objectives need credit-scoring model in place, and the corresponding data-mining objective we would be addressing here is:

Predicting default probability of a customer or borrower, and uncover factors driving default.

7.3.2 Variable Selection for the Model

Since the data did not have any missing value or cleaning issues, data preparation task was reduced to a large extent. Following methods can be used to shortlist the variables to be used in the modeling:

1. Business Sense: Only the variables which were identifiable, actionable are
2. Correlation matrix to avoid using highly correlated data

Chi-square test and Single factor analysis for categorical variables are some other methods that could be used for variable selection.

7.3.3 Variable Transformation

Different variable sets and transformation on them is tried so as to find the best model. These are:

- Categorization vs. non categorization of continuous data
- Different permutation and combinations of variables based on correlation judgment

7.3.4 Modeling

Following modeling techniques are used to model the default prediction in retail banking set up:

1. CART/Decision trees
2. Logistic Regression
3. Neural Networks

Neural networks are kind of representation of a human brain which adapts as more information inflows. It is a multilayer perceptron model that is more of a black box. They are considered to be most accurate, followed by logistic regression and decision trees. However, they are not explainable and hence, their drawback in application. Logistic regression is mostly used in practice as there are hardly any assumptions used other than multicollinearity issue, which we will be taking care of in our approach through correlation testing. The disadvantage of decision tree is its computational burden, and also the fact that it is unstable and even a small change in input data can alter the model considerably.

7.3.5 Evaluation

Various models would be evaluated on the basis of:

- Lift chart/values i.e. improvement in prediction probability of default than in case no model is used
- Applicability in reduction of defaults, insights generated
- Costs/revenues associated

7.4 Findings Through the Modeling Course

The variables provided in data are all actionable in one or the other way unless the customer himself is unknown. However, on constructing the correlation matrix (Appendix A), various high correlation variables (assumed cut-off .35 i.e. top 1% of the outlier correlation value) were found. Such pair of variables along with their correlation values is listed below:

Table 1: Correlations of some highly correlated variables

Variable 1	Variable 2	Correlation
Duration	Amount	.62
History	Cards	.44
Job	Tele	.38
Property	Housing	.35

Hence, one from each set of variables can be rejected so as to avoid multicollinearity issue.

For re-categorization, frequency distribution for each categorical variable was gauged, and clubbed together when distribution too skewed plus it made business sense to merge them. For example, consider the following:

Table 2: Re-categorization Methodology

Duration of Current [redacted] (as given vs. modified)		
Zero	62	234 (<1 year)
< 1 year	172	
<4 years	339	339
<7 years	174	174
>7 years	253	253

SAS enterprise miner has been used to model the default prediction using decision trees, logistic regression and neural networks. Two techniques are applied viz. a model without considering correlation and variable transformation, and one with it.

1.1.1 Modeling with all inputs and no variable transformation

As a first technique, all inputs were fed into the model without any rejection of highly correlated variables and categorization of continuous values. On running all the three models, the results showed high importance given to the highly correlated variables as given in Table 1. This seems to be an issue of multicollinearity. The lift result of this model is as below:

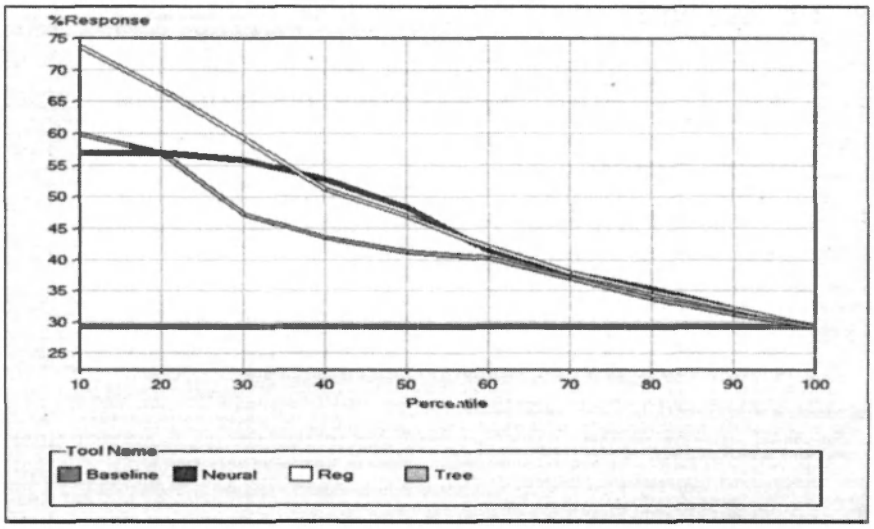

Figure 2 - Lift Chart Without Variable Rejection and Transformation

As can be seen, all models provide a much better prediction rate than the base-line (i.e. when no model is used). When variables were all fed in without any pre-analysis, logistic regression seems to work best. However, if our target base is more than 35% of customers, neural network seems to give better results. But again, due to black box characteristic of neural networks, it doesn't seem to be much effective when application perspective comes in.

Another factor that needs to be considered is cost/profit involved if a customer is misclassified. As an example, the output confusion matrix of decision tree is considered below. Confusion matrix gives the estimation of number of customers who are predicted to default vs. those who actually default, and number of customers who are predicted to not default vs. those who actually not default.

Table 3- Confusion Matrix Without Variable Rejection and Transformation

	Actual default	Actual non-default
Predicted to default	35	52
Predicted to not default	27	186

Without considering the opportunity loss of not offering the loan to people who were predicted to default but did not, we will calculate expected profit per customer using above matrix. Suppose that previously all customers were given loan, and now are given according to above-mentioned matrix. Assuming we gain $1 for a non-defaulting customer and lose $5 for a defaulting one (default loss being much more than profit from lending), the profit from above model is:

Profit over no model $= (186*1 - 27*5) - ((186+52)*1 - (27+35)*5)$

$= 123$

Profit per customer $= 123/300$

$= \$0.41$

1.1.1 Modeling with variable rejection and categorization

The above steps are now followed with all the steps of variable selection, transformation and categorization incorporated. Amount, cards tele, and housing are rejected because of their high correlation with other input variables. Duration and age are categorized into 4 quartiles each. On running the three models, the high importance variables in last model have now moved down the hierarchy as multicollinearity is reduced through rejection of highly correlated variables. The lift results now are as follows:

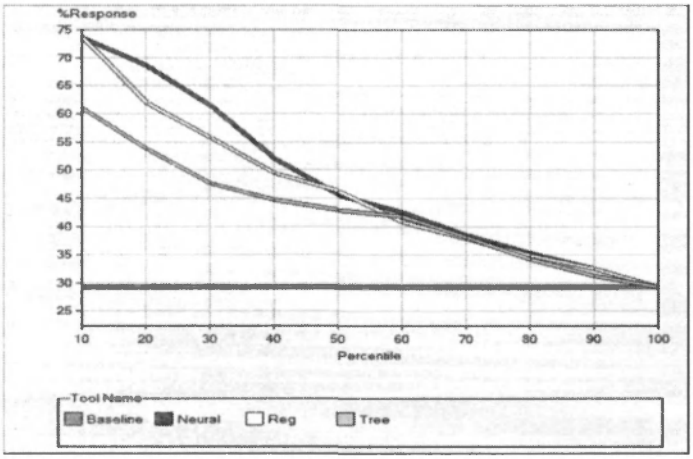

Figure 3- Lift Chart With Variable Rejection and Transformation

All models have improved in their lift performance (73%, 60%, and 55% to 74%, 61%, and 74% at 10 percentile for logistic regression, decision tree, and neural network respectively). The most significant is the improvement in neural network model, which performs better than logistic regression. This result is more in line with the expectations as suggested in various literatures. But again it (neural network) is only helpful in predicting default and not acting upon it once it is known.

Cost profit analysis of the decision tree shows improvement above the previous decision tree model. The confusions matrix in this case is as follows:

Table 4: Confusion Matrix with Variable Rejection

	Actual Default	Actual non-default
Predicted to default	38	62
Predicted to not default	5	195

Taking same assumptions as in previous case, the profit now is as follows:

Profit over no model $= (195*1 - 5*5) - ((195+62)*1 - (5+38)*5)$
$= 128$

Profit per customer $= 128/300$
$= \$ 0.43$

CONCLUSION

It can be seen that even though same modeling technique is used, the efficacy of models depend as much on what variables are used, what transformation are applied, how they are re-categorized, and what business judgments are used. The important variables as found from the above modeling technique are Checking Status, Duration, Age, Other Plans, Marital Status, etc. However, these might differ with availability of more data and from bank to bank. Each bank has to build up its own model so as to predict default and prevent them proactively.

Application of business intelligence and analytics cover almost all departments of banking. Its proper implementation can bring in significant impact on the bottom-line of a bank and also provide with competitive edge. The data mining modeling varies across banks and applications, and hence is not replicable from one to another. Also, the models grow old and are to be reestablished every once in a while. If not done so, it may take a retail bank towards loss rather than gain. The next big thing would be to see how banks utilize big data streams from web, social media, etc. The banks that will be able to identify this opportunity and utilize in effective manner will surely have an advantage over others in the long run.

REFERENCES

Agrawal, R, Dimielinski, T, and Swami, A (1993). Database mining: a performance perspective, IEEE Transactions on Knowledge and Data Engineering 5 (6); 914–925.

Apte, C.V., Hong, S.J., Natarajan, R., Pednault, E.P.D., Tipu, F.A. and Weiss, S.M. (2003). Data-Intensive Analytics for Predictive Modeling, IBM Journal of Research and Development, 47(1): 17–23.

Bartolozzi, E., Cornford, M., Erguin, L.G. and Deocon, P.C. (2008). "Credit Scoring Modelling for Retail Banking Sector", Oscar Ivan Vasquez & Fransico Javier Plaza, II Modeling Week, Universidad Complutense de Madrid, 16th – 24th June, 2008.

Berry, M.J.A, Linoff, G (1997). Data Mining Techniques for Marketing, Sales, and Customer Support, Wiley, New York.

Bose, R. (2009). "Advanced Analytics: Opportunities and Challenges", Industrial Management and Data Systems, 109(2): 155–172.

Brynjolfsson, E., Hitt, L.M. and Kim, H.H. (2011). Strength in numbers: How does data-driven decision making affect firm performance?, Social Science Research Network (SSRN), April 2011.

Budale, D. and Mane, D. (2013). "Predictive Analytics in Retail Banking" International Journal of Engineering and Advanced Technology (IJEAT), Volume 2(5).

Chen; Jason Chou-Hong, P., Pete Chong; and Ye-Sho Chen (2001). Decision Criteria Consolidation: A Theoretical Foundation of Pareto Principle to Porter's Competitive Forces, Journal of Organizational Computing & Electronic Commerce, 11(1); 1-14.

Chitra, K. and Subashini, B. (2013). "Data Mining Techniques and its Applications in Banking Sector", International Journal of Emerging Technology and Advanced Engineering, Volume 3 (8).

Fan, W., Wallace, l., Rich, S. and Zhang, Z. (2006). Tapping the Power of Text Mining, Communications of the ACM, 49(9): 77–82.

Fayyad, U.M., Piatetsky G., Shapiro, Smyth, P and R. Uthurusamy ZEds., (1996). From data mining to knowledge discovery: an overview, in: Advances in Knowledge Discovery and Data Mining, MIT Press, Massachusetts, Chap. 1.

Frawley, W.J., Piatetsky, G., Shapiro, Matheus, C.J. (1992). Knowledge discovery in databases: An overview, AI Magazine 13 (3); 57–70.

Gessner, G.H. and Volonino, L. (2005). Quick Response improves returns of Business Intelligence Investments, Information Systems Management, 22(3): 66–74.

Hofmann, H. "UCI Machine Learning Repository-Statlog" (German Credit Data) Data Set , (link: https://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)

Jain, A. (2014). "Marketing Analytics in Retail Banking" Everest Group Research and Genpact.

Jayasree, V. and Balan, R.V.S. (2013). "A review of Data mining in Banking Sector", American Journal of Sciences 10(10): 1160-1165.

Jourdan, Z., Rainer, R.K. and Marshall, T.E. (2008). Business Intelligence: An Analysis of the Literature, Information Systems Management, 25(2): 121–131.

Jukic, N. (2006). Modeling Strategies and alternatives for Data warehousing Projects, Communications of the ACM, 49(4): 82–91

Kiron, D. et al., (2011). Analytics: The Widening Divide, IBM, MIT Sloan Management Review. Retrieved from: http://sloanreview.mit.edu/feature/achieving-competitive-advantage-throughanalytics

Kocenda, E. and Vojtek, M. (2009). "Default Predictors and credit scoring models for retail banking", Emperical and Theoretical Methods, CESifo working paper no. 2862, Category 12: December 2009.

Kumar, A. and Shenoy, R. (2010). "Analytics in Retail Banking: Why and How?", FINsights- Analytics in Financial Sevices2010 by Infosys Technologies Limited.

Kumar, V., Venkatesan, R., and Rajan, B. (2009). Implementing Profitability through a Customer Lifetime Value Framework, Marketing Intelligence Review, 2, (December): 32-43.

Marketelligent report. "Application of Decision Sciences to Solve Business Problems: Retail Banking Industry".

Michael, H., Kaplan, M.A., Beeser, J.A. (2007). A Model to Determine Customer Lifetime Value in a Retail Banking Context, European Management Journal, 25(3): 221–234.

Pulakkazhy, S., and Balan, R.V.S. (2013). "Data Mining in Banking and its Applications- A Review", Journal of Computer Science 9(10): 1252-1259.

Reinartz and Kumar, V., (2003). The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration, Journal of Marketing, 67 (January): 77-99.

Rust, Roland T., Valarie A. Zeithaml, and Katherine N. Lemon (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy, Journal of Marketing, 68 (January): 23-53.

Sanders, Robert (1987). The Pareto Principle: Its Use and Abuse, Journal of Services Marketing, 1(2); 37-40.

SAS Report, (2012). "Banking on Analytics: How high performance analytics tackle big data challenges in banking".

Shaw, M.J., Subramaniam, C., Tan, G.W. and Welge, M.E. (2001). Knowledge Management and Data Mining for Marketing, Decision Support Systems, 31: 127-137.

Tandulwadikar, A. (2011). "How Analytics can transform the US retail banking sector", Cognizant Research Center – Cognizant Reports, 2011.

Vojtek, M. and Kocenda, E. (2007) "Credit Scoring Methods" (source/elibrary: https://www.risknet.de/)

Wang, H. and Wang, S. (2008). A Knowledge Management Approach to Data Mining Process for Business Intelligence, Industrial Management and Data Systems, 108(5): 622–634.

Watson, H.J. and Wixom, B.H. (2007). The Current State of Business Intelligence, Computer, 40, (9), 96–99.

Watson, H.J. (2013). All about Analytics, International Journal of Business Intelligence Research, 4(1): 13–28.

Weiss, S.M., Indurkhya, N., Zhang, T. and Damerau, F.J. (2005). Text Mining: Predictive Methods for Analyzing Unstructured Information, Springer, New York, NY.

Appendix A: Correlation Matrix

	checkingstatus	duration	history	purpose	amount	savings	employ	installment	status	others	residence	property	age	otherplans	housing	cards	job	liable	tele	foreign
checkingstatus	1.00	-0.07	0.19	-0.06	-0.04	0.22	0.11	-0.01	0.04	-0.13	-0.04	-0.03	0.06	0.05	0.02	0.08	0.04	-0.01	0.07	-0.03
duration		1.00	-0.08	0.11	0.62	0.05	0.06	0.07	0.01	-0.02	0.03	0.30	-0.04	-0.05	0.16	-0.01	0.21	-0.02	0.16	-0.14
history			1.00	-0.03	-0.06	0.04	0.14	0.04	0.04	-0.04	0.06	-0.05	0.11	0.12	0.06	0.44	0.01	0.01	0.05	0.01
purpose				1.00	0.19	-0.02	-0.03	-0.03	-0.02	0.08	0.04	0.08	0.04	-0.15	0.08	0.02	0.10	0.00	0.12	0.02
amount					1.00	0.06	-0.01	-0.27	-0.02	-0.03	0.03	0.31	0.03	-0.05	0.11	0.02	0.29	0.02	0.28	-0.05
savings						1.00	0.12	0.02	0.02	-0.11	0.09	0.02	0.08	0.00	0.01	-0.02	0.01	0.05	0.09	0.01
employ							1.00	0.13	0.11	-0.01	0.25	0.09	0.26	-0.04	0.11	0.13	0.10	0.10	0.06	-0.03
installment								1.00	0.12	-0.01	0.05	0.05	0.06	0.00	0.09	0.02	0.10	-0.07	0.01	-0.09
status									1.00	0.05	-0.03	-0.01	0.01	-0.04	0.10	0.06	-0.01	0.12	0.03	0.07
others										1.00	-0.03	-0.16	-0.03	-0.06	-0.07	-0.03	-0.06	0.02	-0.08	0.12
residence											1.00	0.15	0.27	0.00	0.01	0.09	0.01	0.04	0.10	-0.05
property												1.00	0.07	-0.09	0.31	-0.01	0.28	0.01	0.20	-0.13
age													1.00	-0.04	0.30	0.15	0.02	0.12	0.15	-0.01
otherplans														1.00	-0.07	-0.05	0.00	-0.08	-0.02	0.02
housing															1.00	0.05	0.11	0.11	0.10	-0.06
cards																1.00	-0.03	0.11	0.07	-0.01
job																	1.00	-0.09	0.38	-0.10
liable																		1.00	-0.01	0.08
tele																			1.00	-0.11
foreign																				1.00