DATA MINING: A SOLUTION FOR DATA MANAGEMENT IN THE BANKING SECTOR

Rakesh Kumar Giri
Assistant Professor,
Bharath University, Chennai, India

Abstract: The information technology revolution is one of humanity's most significant achievements. The usage of information technology is very common in all enterprises, regardless of size. The banking industry is one of the sectors that has incorporated the use of information technology into all of its activities. Banks' data bases are growing at such a rapid pace that they are unable to manage them and so cannot use the data for proper decision making. Some of the world's leading banks have begun to use cutting-edge data mining techniques for customer segmentation, profiling, fraud detection, targeting new customers, retaining existing customers, and predicting future trends, while others are still planning their implementation. The primary goal of this article is to analyze the challenges encountered by banks when handling data and to determine whether banks that use data mining technologies perform better in this regard than banks that do not use data mining techniques.

Index Terms **- Data Mining, Decision Making, Fraud Detection, Customer Relationship Management**

I. INTRODUCTION

Banking sector around the globe is facing a paradigm shift. The whole concept of banking has been shifted to centralized databases, online transactions, network connectivity and ATM's all over the world, which has made the banking system technically strong and more customer oriented. The huge size of these data bases makes it impossible for the organizations to analyze these data bases as per the need of the decision makers [2]. Since 1980's the Indian banking sector is incorporating the concept of Management Information System, through which banks are generating various kinds of reports, which are analyzed for the decision making within the organization [1]. The Total Branch Computerization (TBC) software packages being used at various branch levels are transaction oriented, as these were designed keeping day to day transactions in mind. Designing the new MIS or restructuring the existing ones would not be possible as such. The solution seems to be in incorporating the concept of data warehousing and data mining. The concept of data warehouse is coming to a reality only due to advances in the techniques of data capturing; processing, transmission, and storage. Moving a step further, to extract, analyze, project and mine the data as per the requirements of the users and analysts, the data mining technology is helping the organization. It is the process adopted to undertake a thorough analysis of the data available to the firm to select the information, identify patterns and relationships amongst data and to make this information available in the right form for making the decision-making process more effective and easy. Data Mining is developed with the goal of providing tools for automatically and intelligently transforming large amount of data in to knowledge relevant to the users. The extracted knowledge is often expressed in the form of association rules decision tress or clusters that allows one to find interesting patterns buried deeply in the data which facilitate decision making process.

Data mining is defined as the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions.
Cabena et al. Data mining is described as the automated analysis of large amounts of data to find patterns and trends that may have otherwise gone undiscovered.

The banks are facing the following problems in handling and managing the data:

(a). Too much data, but little information is available. (b). Which set of customers is loyal and disloyal?

(c). Whether the loan sanctioned by the bank will be recovered or not i.e to find the right set of customers for providing loans.

(d). Scanning the churn behavior of the customers and then to design strategies to prevent it and hence to retain the valuable customers.

(e). Detection of the transactions leading to fraud.

(f). Customization of services as per the requirements of the users.

(g). Cross selling the products to existing customers.

(h). Which segment of population is most likely to respond to a particular service?

(i). Whether two market segments have similar characteristics?

(j). Analysis of the past trends to determine the present demand and to forecast the customer behavior of various products and services in order to grab more business opportunities and anticipate behavior patterns in the future.

II. LITERATURE REVIEW

Data mining is capable of answering questions about the past (what has happened), the present (what is happening), and the future (what might happen) - (Nemati &Barko, 2002) [3] Keeping the requirement of use of information technology in the banking sector, the Reserve Bank of India constituted a committee on technology upgradation in the banking sector (1999)[5], the committee highlighted the usage of management information systems by the banks and recommended that all banks have to use information technology for the various operations performed by them and should report in electronic formats only and the committee recommended the use of data warehouse and data mining techniques. According to HDFC Bank's Vice-President (Retail Marketing), Mudit Saxena[6]: ``Data mining helps to increase sales by targeting the right customers and to make the right offers to customers. Banks, who have their ears to the ground regarding their customer's tastes and preferences, gather a lot of data...what data mining does is that it sifts through all the voluminous data and ekes out a pattern, which enables the bank to personalize its communication towards the customer as much as possible. With data mining, banks can get a better understanding of the facts which are responsible for better customer relationship. Dean P. Foster, Robert A. Stine(2004)[7] explained that bankruptcy can be predicted by using standard regression techniques to large, complex data sets by making a few assumptions about the structure of the data. The personal bankruptcy was predicted among users of credit cards. This application presented many challenges, ranging from the rare frequency of bankruptcy to the size of the available database. Rajanish Dass (2006)[8] suggested that data mining techniques can be of immense help to the banks and financial institutions in this arena for better targeting and acquiring new customers, fraud detection in real time, providing segment based products for better targeting the customers, analysis of the customers' purchase patterns over time for better retention and relationship, detection of emerging trends to take proactive approach in a highly competitive market adding a lot more value to existing products and services and launching of new product and service bundles.

Data mining can improve the response rates in the direct mail campaigns as the time required to classify the customers will be reduced, this in turn will increase the revenues and will improve the sales force efficiency. M Purna Chandar, Arijit Laha, P Radha Krishna (2006)[9] opines that acquiring new customers is costlier process than retaining existing customers. So managing the customers and maintaining good relationships with them is of utmost importance. The prediction of churn behavior, target marketing, cross/up selling, customer profiling, etc are very important for all the organizations. In Indian banking system, the banks have not arranged their data earlier as per the requirements for prediction of their churn behavior, so the raw customer data should be converted into meaningful data that suits modeling churn behavior. Then, this meaningful data can be converted into knowledge by using

predictive data mining techniques. Chowdari Prasad(2007)[10] opined that the concept of banking has been transformed from commercial Banking to convenience banking. Banking Industry in India has witnessed a sea change as per the changing customers' needs and regulations from governing authorities. The function of a bank which was only accepting the deposits and lending on selective basis has been changed to relationship banking, private banking, etc. Today, healthy competition has set in between public, private and foreign banks – all of them are offering a wide variety of products and services. Legal and Technological changes have made the banks to take care of the tech-savvy customers, so that they can use Electronic Banking,

Internet Banking, Retail Banking, Online Banking, Mobile Banking or Virtual Banking at par with banks all over the world. Madan Lal Bhasin (2006)[11] opined that the leading banks are using data mining tools for customer segmentation and profitability, credit scoring and approval, predicting payment default, marketing, detecting fraudulent transactions, etc. Chase Manhattan Bank in New York, was facing a financial crunch mainly due to constant decrease in the customer base, then the bank used the techniques of data mining to analyze customer profiles and to use them for their benefits and hence chalked out the strategy for the survival and succeed in its attempt. Data mining is also being used by Fleet Bank, Boston, to identify the best candidates for mutual fund offerings.

III. OBJECTIVES OF THE STUDY

Data mining can play an important role in a decision-making system. It provides a methodology for problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning and innovations. - Hedelin & Allwood, 2002, Liao, 2003 [4]. The objective of the study is to find out whether the use of data mining techniques can be helpful in the banking sector or not.

IV. RESEARCH METHODOLOGY

The primary data has been collected from employees of public and private sector banks. The questions asked from the respondents and their responses have been shown in Annexure I and apart from the technical questions, the demographic details collected are: Gender, Age, Experience in Banking Sector, Experience of Using Any Software Package, Primary Job Function i.e Marketing, Operations or IT. The sample has been selected using Stratified Random Sampling Technique. The population has been divided into three categories i.e. employees working at senior, middle and junior levels and then this set of employees have been further divided on the basis of their area of operations that is marketing, operations or information technology. 100 respondents each from public and private sector banks have been selected. For the purpose of collecting data 100 respondents from Punjab and 100 respondents from Haryana have been considered. It was further studied that among 200 genuine respondents 80 respondents i.e. 40% respondents are from those banks where already data mining concept is in use and the rest 120 respondents are from the banks which are not using the data mining concept in their banks.

The list of banks included in the survey is:

PUBLIC SECTOR BANKS	PRIVATE SECTOR BANKS
• Punjab National Bank	• HDFC Bank • ICICI Bank
• State Bank of Patiala • State Bank of India • Oriental Bank of Commerce	• IDBI Bank • Axis Bank

V. RESULTS & ANALYSIS

The objective of the study is to investigate the major problems faced by the banks while handling the huge amount of data in the banks and whether the banks which are using data mining techniques are in a better position to handle data and hence are better in making strategic decisions regarding the customers and planning the various activities.

This section analyzes the problems faced by the banks while handling the huge amount of data. As the competition is increasing around the globe, the banking sector is also feeling the heat. Normally the banks are not able to retrieve the information required for the current analysis; banks are flooded with data but still striving for the knowledge. Another major point is that the information required for the decision making purposes is need not to be gathered through extensive surveys, but it is already available in the data bases of the bank. The need is to extract the relevant data and to convert it into the form appropriate for the decision making process.

It has also been observed from the collected data, that these problems are not faced or are faced to lesser extent by the banks which are using data mining techniques. Comparative analysis has been used to analyze the degree of problems as it is measured on 5 – point Likert reversed scale (strongly disagree to strongly agree) among the banks which are using data mining software in their daily routine and among those who are not using these techniques.

The table 1.1 shown in Annexure I depict the frequency distribution pattern of responses and comparison of distribution of responses among the banks classified on the usage of data mining software's.

The analysis given in table 1.2 shown in Annexure I shows that, on issues like difficulty to retrieve data from data base as per the requirement, the respondents from the banks which do not use any data mining software, 58% respondents agree and there are 37% respondents who are confused on this issue whereas the respondents from banks where data mining software is in use in totality 92% respondents disagree that means they do not face such problems.

The banks are storing huge amount of data, still they are striving for knowledge, When respondents were asked their opinion on this issue, there is a contradiction in the opinion of the respondents from the banks which are using data mining and who are not using this technique, taking the case of banks which do not use data mining, about 49% respondents complain that too much data is present but they have little information and 32% respondents among them are again confused on this issue and on contrary to this 88% respondents do not agree on this issue as they are using data mining software in their banks.

The respondents from banks where data mining software is in use, 84% respondents reported that it is not difficult to differentiate between loyal and disloyal customers while 57% respondents from banks where data mining software is not in use feel difficult to segment the customers between loyal and disloyal, while among same group 24% respondents have disagreed on this issue.

The occurrence of frauds in banking transactions is very common these days. The banks are able to track such transactions are not, when opinion was asked on this issue i.e The differentiation of transactions leading to fraud is difficult, is being accepted by 63% respondents from those banks where data mining software is not in use and among them 28% are neutral and 9% respondents disagree on this issue, while the respondents from banks where data mining software is in use approx 70% respondents do not agree to it and rest 29% respondents are neutral on this issue. The clustering technique of data mining can be helpful here to group the similar entities together and hence the banks can deal with the different groups separately. On the issue of providing customer oriented services is quite difficult is being accepted by 57% respondents from the banks where data mining software is not in use while 84% respondents feel disagreed on this issue as they are using data mining software in their banks and among both categories respondents approx 24% and 16% respondents respectively are neutral.

The 74% respondents have disagreed that conversion rate of call center executive is very low, as they are using data mining software in their banks and rest 26% respondents from same category are neutral on this issue. The respondents from the banks where data mining software is not used, approx 54% respondents think conversion rate of calls is low while 37% respondents from same group are neutral and rest 10% have disagreed.

67% respondents from those banks where data mining software is in use disagree on the issue that it was difficult to judge the tastes and preferences of the customers while 31% respondents from same group are neutral. On the same issue among those respondents where data mining software is not in use, 52% of the respondents agree on this issue and 42% are neutral.

There exists no exact method to predict that customer would pay his loan or not, on this issue the respondents from the banks where data mining software is in use 71% respondents disagree on this issue, while the remaining 29% of the respondents are neutral. The group of the respondents from the banks where data mining software is not in use, 48% respondents agree on this issue while 20% respondents disagree and 32% respondents are neutral.

Current software's are not able to provide the relevant information as per the requirements of the decision makers, on this issue the respondents from the banks where data mining software is in use 77% respondents disagree while 21% of the respondents are neutral. The group of the respondents from the banks where data mining software is not in use, approx 60% respondents agree on this issue while 31% respondents disagree and 9% are neutral respondents.
Also statistically it has been analyzed that there is a significant difference in average response of the respondents regarding allproblems, categorized by the banks where data mining software is used or not. The null hypothesis used to analyze the significant differences is, H_0: There is no significant difference between the average response level of the respondents for all the problems mentioned, categorized by the banks where data mining software is in used or not.

The analysis of the table 1.2 show that all of the discussed problems in this section related to problems in handling data, there is significant difference between the response levels of the respondents from the banks where data mining software is used and among those banks where this is not used. The null hypothesis is rejected as all p – va lues obtained from Mann Whitney statistics are less than 0.05.
Also, it is analyzed that the average response for all problems discussed from respondents from the banks where there is no usage of data mining softwares is in the range of 3.29 – 3.54 which is much more than the average res ponse of those respondents from the banks where data mining softwares is in use i.e. 1.71 – 2.31.

Since the reversed Likert scale was used i.e. strongly disagree to strongly agree, thus the response level of the respondents from the banks where data mining software's is in usage falls in the range of strongly disagree – disagree – neutral whereas the response level of the respondents from the banks where data mining software's are not in usage falls in the range of agreed on all problems.

VI. DISCUSSIONS AND CONCLUSION

The usage of information technology is proving to be a boon for the banking industry. The need of latest techniques and tools is being felt to increase the overall efficiency of the system. The ever growing competition is forcing the banks to adopt the latest techniques to stay ahead of others. Data mining is one of the latest techniques which can help the banks for providing better customized services to customers, detecting frauds, forecasting future trends and to help the decision makers for making better decisions. In India, the banks have realized the need of such techniques and the study conducted here reveals the fact that the banks implementing these techniques are in a better position to handle the problems related to data management and decision making. The banks which have not implemented these techniques are facing the problem in acquiring new customers, retaining existing customers, detecting frauds, detecting and preventing attrition of the customers and hence a strong need is felt by the banks to overcome these mentioned problems. The study may put forward as a guide for the banks who have not thought of implementing data mining techniques or are still planning to implement these techniques. Such organizations can fasten the process of implementation of these techniques. The need is to create the awareness about the prospective benefits of data mining technology to the management, so that the organizations can get the full benefits by the proper and timely implementation of this technique.

REFERENCES

[1] Danziger, J. N., & Andersen, K. V. (2002). 'The Impacts of Information Technology on Public Administration: An Analysis of Empirical Research From The 'Golden Age' of Transformation'. International Journal of Public Administration, 25(5), 591-627.

[2] Engler, H., & Essinger, J. (2000). The future of banking. UK: Reuters, Pearson Education.

[3] H. R. Nemati and C. D. Barko(2003)"Key Factors for Achieving Organizational Data Mining Success". Industrial Management and Data Systems, vol. 103, no. 4, pp. 282–292.

[4] Hedelin, L., & Allwood, C. M. (2002). 'IT and strategic decision making'. Industrial Management + Data Systems, 102(3/4), 125.

[5] A. Vasudevan(1999) Report of the committee on technology up-gradation in banking sector" Reserve Bank of India, May 1999, chapter 6.

[6]. Mudit Saxena(2000), Vice-President (Retail Marketing), HDFC Bank, Press Statement, Indian Express Newspaper on 9th Nov, 2000.

[7} Foster, D. P. and Stine, R. A(2004) "Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy". Journal of the American Statistical Association, Alexandria,Vol. 99, pp. 303-313.

[8] Rajanish Dass (2006), Data Mining In Banking And Finance: A Note For Bankers- Technical note, Note No.: CISG88., April 2006

[9]M. Purna Chandar, Arijit Laha and P. Radha Krishna Modeling Churn behavior of bank Customers using Predictive Data Mining Techniques", National Conference on Soft Computing Techniques for Engineering Applications, (SCT−2006), Published by Institute for Development & Research in Banking Technology [IDRBT]

[10]Chowdari Prasad(2007), Commercial Banking to Convenience Banking. Faculty Coloumn, IndianMBA.com.

[11]Madan Lal Bhasin (2006), "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October, 2006 .

Annexure I

Table 1.1 Response Pattern for Problems in Handling Data (All Respondents)

	Usage of Data Mining (Yes/No)	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Sometim findin es g the data from the data base, as per the requirement is difficult.	Yes	36.9%	55%	8.2%	-	-
	No	-	4%	37.6%	58.4%	-
Too much of data is present but little information is available.	Yes	26.5%	61.6%	11.9%	-	-
	No	-	17.8%	32.7%	49.5%	--
Segmentation of customers between loyal and disloyal is difficult.	Yes	20%	64.4%	15.6%	-	-
	No	-	23.8%	18.8%	57.4%	-
Differentiation of transacti ons leading to fraud is difficult.	Yes	12.9%	56.2%	29.2%	-	
	No	-	8.9%	27.7%	63.4%	--

Providing customer oriented services is quite difficult.	Yes	20%	64.4%	15.6%	-	-
	No	-	18.8%	23.8%	57.4%	-
Conversion rate	Yes	6.4%	67.1%	26.5%	-	-
Phone calls by call centre executive is very low.	No	-	9.9 %	36.6%	53.5%	-
It is difficult to judge the taste and preferences of the customers.	Yes	5.9%	61.1%	31%	1%	-
	No	-	5%	42.6%	52.5%	-
There exists no exact method to predict whether the customer will be able to repay the loan or not.	Yes	15.1%	55.4%	29.5%	-	-
	No	-	19.8%	31.7%	48.5%	-
Current softwares are not able to provide the relevant information as per requirements of the decision makers.	Yes	22%	55.4%	21.3%	1.2%	-
	No	-	8.9 %	30.7%	60.4%	-

Table 1.2 Analysis of Responses for the Problems in Handling Data
(Dependency on Usage of Data Mining Techniques)

	Responses from the Banks where Data Mining is not used.				Responses from the Banks where Data Mining is used.				Mann – Whitney Test
	Mean	Median	Min	Max	Mean	Median	Min	Max	
Q1	3.54	4.00	D	A	1.71	2.00	SD	N	0.00
Q2	3.32	3.00	D	A	1.85	2.00	SD	N	0.00
Q3	3.34	4.00	D	A	1.96	2.00	SD	N	0.00
Q4	3.54	4.00	D	A	2.20	2.00	SD	A	0.00
Q5	3.39	4.00	D	A	1.96	2.00	SD	N	0.00
Q6	3.44	4.00	D	A	2.20	2.00	SD	N	0.00
Q7	3.44	3.00	D	A	2.31	2.00	SD	A	0.00
Q8	3.29	3.00	D	A	2.14	2.00	SD	N	0.00
Q9	3.51	4.00	D	A	2.02	2.00	SD	A	0.00