A Study on Role of Data Mining for Customer Relationship Management in Banking Sector

Mr. Sushim Kumar, MBA-15 (2018-20)[1]
Prof SubhranshuMohanty, Assistant Professor (IT)[2]

Abstract

Data mining is one of the tasks in the process of knowledge discovery from the database. Inthe corporate world every organization is competing with the other organization in terms oftheir value towards the business and the financial growth. Apart from execution of the businessprocesses, the creation of knowledge base and its utilization for the benefit of the organizationis becoming a strategy tool to compete. In this research, researcher discuss about the basicdetails of data mining, the use of knowledge discovery process and the new techniques fromthe business point of view in the banking Industry.

The growth of the organization dependson the quality of service, competing with the other organizations, provide required informationto the customers, satisfaction of the employees working in the organization. In the bankingsector all the financial work can be done in the computers and their connectivity through WorldWide Web. The software's get automatically updated in time, use of internet banking and ATMhave made a big change in the banking sector.

The banks have realized that their biggest asset isthe knowledge and the planning to implement the right knowledge at the right time in the right market.

Keywords: Data Mining, Knowledge Discovery in Databases (KDD), Knowledge Base.

BACKGROUND OF THE STUDY

In the past, all companies have focused on selling products and services without having or looking for knowledge about their own customers, but with increasing competition, attracting new customers has become more difficult and in this context, companies need to change their approach and step up their efforts to keep current customers. Social and economic evolution has changed consumer lifestyles, customers have become better informed and less inclined to respond to marketing communications.

Thus, this change in consumers' behaviour has led companies to evolve, change their approach and have customer-oriented strategies, constantly improve service quality to ensure a good business relationship with customers. Companies have Customer Relationship Management departments specifically to make strategies to increase customer retention and development. Banks also needed to realize that customer relationships are very important for sustainable growth in the long run and Customer Relationship Management (CRM) is the strategy that can help build these long-term relationships and implicitly increase its revenue and profits. Developing the IT sector has helped the banking industry store stunning amounts of customer data, analyse and

[1]Student, MBA-15 (2018-20)
[2]Assistant Professor (IT)
Army Institute of Management and Technology, Greater Noida

interpret this data in order to develop marketing strategies, risk, but also to improve processes and to better understand customers with the help of data mining.

INTRODUCTION

Data mining is becoming strategically important area for many business organizations including banking sector. It is a process of analysing the data from various perspectives and summarizing it into valuable information. Data mining assists the bank to look for hidden pattern in a group and discover unknown relationship in the data. Today, customers have so many opinions with regard to where they can choose to do their business. Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations for the banking sector to avoid customer attrition. Customer retention is the most important factor to be analysed in today's competitive business environment. And also fraud is a significant problem in banking sector. Detecting and preventing fraud is difficult, because fraudsters develop new schemes all the time, and the schemes grow more and more sophisticated to elude easy detection. This research analyses the use of data mining techniques and its applications for customer relationship management in banking sector like fraud prevention and detection, customer retention, marketing and risk management.

With the development of technology, the infamous Internet services, competitive process by increasing access to customers according to different approaches to management companies in connection withthe customers' show different companies for survival in market competition wished more profit in business need to analyse that in the market will be done. In fact, that the future oriented companies were identified and needs of their own customers' infection more belief and showed their own customers' infection is considered a value for them. Always in competitive markets more successful action. The companies via information in connection with their own customers' infection models that high–ranking bought from other rivals. Infamoustheir products and services with regard to this information.

DATA WAREHOUSING AND DATA MINING

DATA WAREHOUSE

Data Warehouse (DW), also known as an Enterprise Data Warehouse (EDW), is a system used for reporting, analysing data which is considered a core component of business intelligence, Dedic N. and Stanier C.(2016). Data Warehouses are central repositories of integrated data from one or more disparate sources. They store current and historical data in one single place that are used for creating analytical reports for knowledge workers throughout the enterprise.

The data stored in the warehouse is uploaded from the operational systems (such as marketing or sales). The data may pass through an operational data store and may require data cleansing for additional operations to ensure data quality before it is used in the DW for reporting.

There is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyse this huge amount of data andextract useful information from it. Data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once allthese processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc.

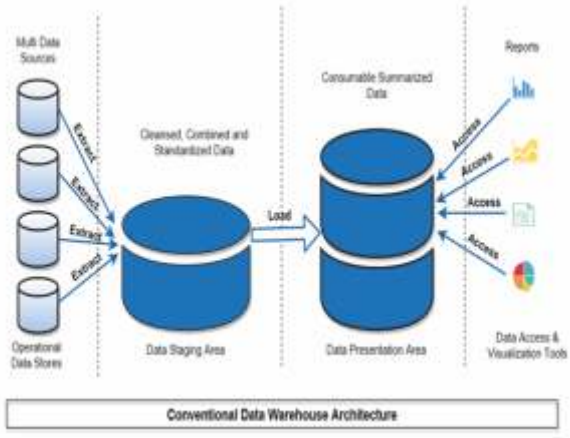

Fig.1: Architecture of Enterprise Data warehousing
Source: https://www.teksystems.com/

- Data Sources (operational systems and flat files)
- Staging Area (where data sources go before the warehouse)
- Warehouse (metadata, summary data, and raw data)
- Data Marts (purchasing, sales, and inventory)
- Users (analysis, reporting, and mining)

DATA MINING

With time and increasing data volume and expansion of modern data storage size, nature and management of statistical methods, Data necessary to evaluate the extraction of knowledge from the show. The best solution to meet this data mining can be cited. We show in Table under evolutionary data mining techniques to collect data.

Technology	Evolution
Disks of computers	data collection(1960)
Relational database(ODBC, SQL)	access data(1980)
OLAP (Online analytical processing)	data warehouse DSS(1990)
Multi-processing computer professional	data mining(2000)

Table 1.1: The evolution of data collection methods

Data mining process to extract information and advanced data analysis, and extracting knowledge from massive amounts of data in a database format, a data warehouse or any information storage is saved. In fact very similar to the data mining is the extraction of metals from the mountain. The data warehouse and data mining information that lies in the extract.

Basis of a data mining process involves five stages as follows:

- A set of training samples should be chosen, are collected and trimming.
- Type of knowledge: knowledge of the expected data mining techniques used will be specified.

- Knowledge base: the transfer of existing knowledge about the process
- Evaluation criteria: criteria value of knowledge gained from data mining, extraction time and knowledge in what has been the representation of key importance and will help data mining process.
- Presentation: usually extracted is determined depending on the type of knowledge. In many cases there is also suitable for representation.

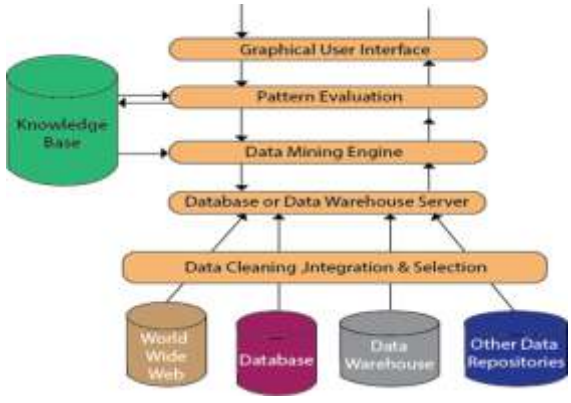

Fig.2: Architecture of Data Mining
Source: https://www.javatpoint.com/

Data mining activities toward knowledge discovery includes the following steps:

1) Identify the purpose and scope of its application is clear understanding that what, in what will be done within a field.
2) Select the data analysis and discovery for purposes of determining the preparation of data, including data cleansing
3) adopt the best methods for achieving the goals of data mining
4) The application of data mining algorithms
5) Evaluation and validation results
6) The use of stabilization and consolidation of results and knowledge discovered making decisions based on the knowledge discovered

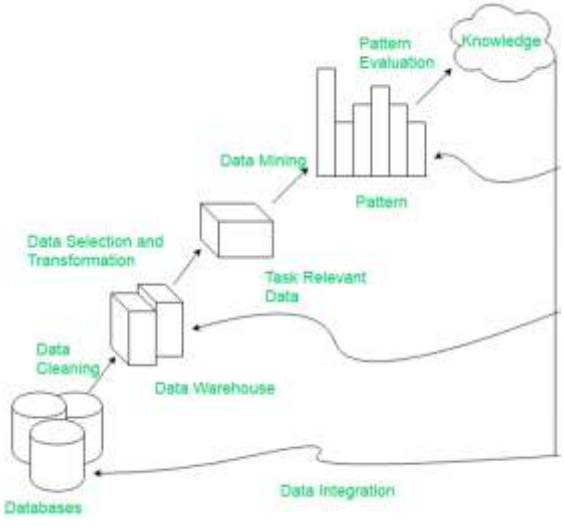

Fig. 3:KDD (Knowledge Discovery in Database) Process
Source: https://tutorialspoint.dev/

Four basic approaches to data mining has the following:

- Classification
- Regression
- Clustering
- Association

Classification: Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we make the software that can learn how to classify the data items into groups.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

Regression: Technique used to predict a range of numeric values (also called continuous values), given a particular dataset. For example, regression might be used to predict the cost of a product or service, given other variables.

Regression is used across multiple industries for business and marketing planning, financial forecasting, environmental modelling and analysis of trends.

Clustering: Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as pre-processing approach for attribute subset selection and classification.

Types of clustering methods:

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

Association: Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. Association and correlation is usually used to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behaviour analysis.

The various types of associations include:

- Multilevel association rule.
- Multidimensional association rule.

- Quantitative association rule.
- Direct association rule.

CUSTOMER RELATIONSHIP MANAGEMENT (CRM)

Dimensions of Customer Relationship Management.

- Customer Identification
- Customer Attraction
- Customer Retention
- Customer Development

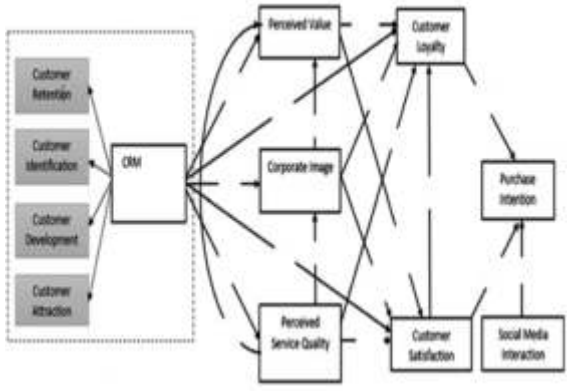

Fig. 4: CRM Components (Kracklauer, Mills and Seifert, 2004)
Source: http://www.utgjiu.ro/

Like any other business process, CRM follows a life cycle, and hence it is necessary to recognize the importance of each of the various stages in the life cycle for CRM to succeed.

Hyperion Solutions proposed a model to determine the life cycle of CRM processes.

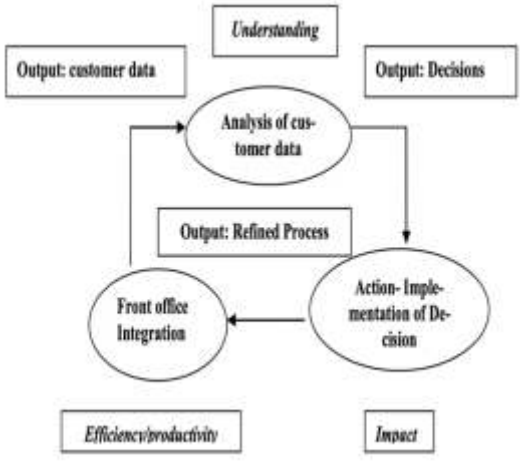

Fig. 5: CRM Life Cycle
Source: http://www.hsg-llc.net/

This model of CRM life cycle comprises of three stages:

Stage I, the integration stage, consists of integrating the front office systems, and centralizing the customer data. While the output of this stage is a centralized source of all customer data, its benefits include Efficiency and productivity.

Stage II, the life cycle is the analysis stage where customer data are analysed to understand the patterns in customer behaviour. This helps in making decisions about the various strategies to be implemented in order to increase customer profitability. It is in this stage that data mining plays a major role in understanding and predicting the customer behaviour.

In the final stage, Action, the decisions are implemented which will have an impact on the business and organizational processes. The outcome of this stage is the refinement of these processes based on the improved customer understanding gained through analysis.

DATA MINING FOR CRM

The important role in the process of data mining "CRM" plays. Because on one hand, with data canters or data warehouse is the interaction and the interaction with the software analysis is competitive management. Relationship between the manual management competitive analysis software with data mining software requires the transcription of data models to the data transmission. The same definition in the software sector, customer data mining and analysis software for corporate management to ensure that competitive. The model need not have your entire database.

Data mining in customer relationship management process is shown in figure 6.

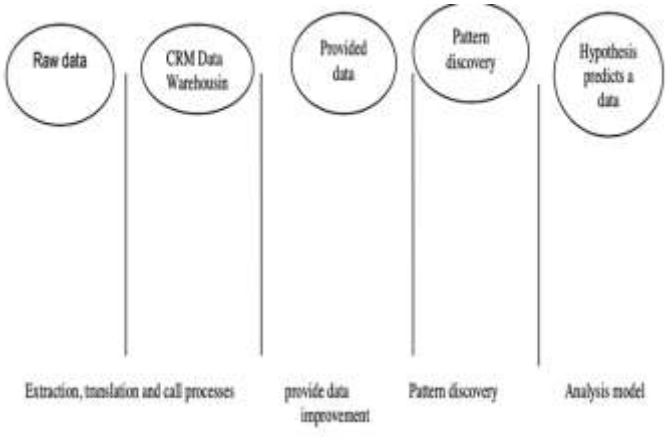

Fig. 6: Data Mining process in CRM
Source: https://www.datasciencecentral.com/

Raw data are collected from various sources and the extraction, translation and management processes of this type are called data warehouse. In data preparation, data out of the warehouse and data mining are an appropriate format.

The overall advantages of using "data mining" in "CRM":

- Data filtering to eliminate duplicate data.
- Extraction, data management, analysis and access to utility customers, retain customers models.
- Rapid and accurate access to integrated data.
- The use of precision instruments and advanced data analysis and reporting.
- Increase customer satisfaction.
- To attract potential customers, retain existing customers and increase market share.

In total, over a period of data mining can be used to predict the exact changes. Organizations in order to fetch data to identify patterns in data mining helps customers.

An Integrated model of Data mining for CRM:

It is based on inputs from Parsaye's classification of Data mining processes and CRM Group's model of the CRM Process.

The integrated model consists of the three types of data mining processes, three stages of the CRM process, certain specific techniques of data mining that could be used in the various stages of the CRM process and the possible applications of those techniques.

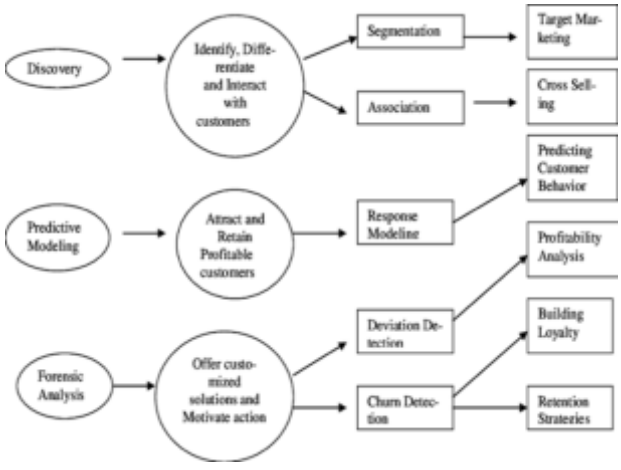

Fig. 7:Integrated model of Data Mining for CRM
Source: https://www.researchgate.net/

The applications of the data mining techniques become the important triggers for making strategic decisions related to the Customer Relationship Management process. The implementation of these actions is carried out in the action stage of the CRM life cycle.

RESEARCH OBJECTIVES

The objectives of this research are as follows:

- To analyse the data mining applications that can be helpfulfor improving customer relationship management.
- To study the major technological requirements of CRM system that support building data mining applications.
- To illustrate the contribution of data mining in CRM.

LITERATURE REVIEW

Constantin BrâncuşI (2018) The aim of this paper is to present the concept of data mining and the concept of data discovery (KDD), but also the impact and important use of data mining techniques in the banking sector. This paper explores and reviews various data mining techniques that are applied in the banking sector but also provides insight into how these techniques are used in different areas to make decision-making easier and more efficient.

Hossein Hassani, XuHuang and Emmanuel Silva (2018) their paper contributed to bringing valuable insights with regard to the future developments of both DM and the banking sector along with a comprehensive one stop reference table. Moreover, they identified the key obstacles and present a summary for all interested parties that are facing the challenges of big data.

Yuvika Priyadarshini (2017) The aim of the study was to identify the extent of Data mining activities that are practiced by banks. Data mining is the ability to link structured and unstructured information with the changing rules by which people apply it. It is not a technology, but a solution that applies information technologies. Currently several industries including like banking, finance, retail, insurance, publicity, database marketing, sales predict, etc are Data Mining tools for Customer.

M. Preethi, M.Vijayalakshmi (2017) Their article analysed the various data mining techniques and concepts that can be applied to banking sector to enhance its performance.

Burcu Oralhan, Kumru Uyar, ZekiOralhan (2016) Their aim was to investigate effect of 6 factors on customer churn prediction via data mining methods. After sale service software database is the source of our data. Their data source variables were Customer Type, Usage Type, Churn Reason, Subscriber Period and Tariff. The data was examined by data mining program. Data was compared in 8 classification algorithm and clustered by simple K means method. They determined the most effective variables on customer churn prediction.

Ahmet Boyaci, Kasım Baynal, Aslı Calis (2015) The study conducted in banking sector, it was aimed to reduce the rate of risk in decision making to a minimum via analysis of existing personal loan customers and estimate potential customers' payment performances with k-means method is one of the clustering techniques and the decision trees method which is one of the models of classification in data mining. In their study, SPSS Clementine was used as a software of data mining and an application was done for evaluation of personal loan customers.

Sandeep Sharma, Sashi Tarun, Pankaj Sharma (2014) have discussed about the basic details of data mining and the use of knowledge discovery process and the new techniques from the business point of view. In their approach they made an efficient system so that the organization will get the right information at the right time and right to access the necessary information for their growth. The growth of the organization depends on the quality of service, competing with the other organizations, provide required information to the customers, satisfaction of the employees working in the organization.

Dr. K. Chitra, B. Subashini (2013) Their paper analyses the data mining techniques and its applications in banking sector like fraud prevention and detection, customer retention, marketing and risk management.

RESEARCH METHODOLOGY

In a view to precede the research in a systematic way the following research methodology has been used. As the nature of research in CRM and data mining are difficult to confine to specific disciplines, the relevant materials are scattered across various journals. The methodology adopted in the study is both descriptive and analytical. Secondary data was collected from different official bank websites and the following online journal databases were searched to provide a comprehensive bibliography of the academic literature on CRM and Data Mining:

a. ABI/INFORM Database
b. Ebsco
c. Academic Search Premier
d. Research Gate
e. Business Source Premier
f. Emerald Fulltext
g. Ingenta Journals
h. Science Direct
i. IEEE Transaction

DATA ANALYSIS AND INTERPRETATION

Each of the selected articles was reviewed and classified according to the proposedframework. The classification process is consisting of four phases:

1) Online database search.
2) Initial classification by researcher.
3) Independent verification of classification results by any secondresearcher.
4) Final verification of classification results by third researcher.

If there was a discrepancy in classification, each of these articles was then discussed until there was agreement on how the article should be classified from the final set in the proposed classification framework. The selection criteria and evaluation framework is shown in Fig. 8. The collection of articles was analysed in accordance with CRM elements and data mining models, by year of publication and according to the journal in which the article was published.

Fig. 8: Selection criteria and evaluation framework
Source: Secondary data

TABLE NO: 5.1

DISTRIBUTION OF BANKS BY CRM AND DATA MINING MODEL

CRM dimensions	CRM elements	Data Mining Model	Number of Banks		
Customer Identification	Customer segmentation		4		
		Classification		1	
		Clustering		2	
		Regression		1	
	Target customer analysis		4		
		Classification		2	
		Clustering		1	
		Visualization		1	
					8
Customer Attraction	Direct marketing		5		
		Regression		1	
		Classification		3	
		Clustering		1	
					5
Customer Retention	Complaints management		2		
		Clustering		1	
		Sequence		1	
		Discovery			
	Loyalty program		8		
		Classification		4	
		Clustering		1	
		Regression		2	
		Sequence		1	
		Discovery			
	One to one marketing		9		
		Association		3	
		Classification		1	
		Clustering		3	
		Sequence		1	
		Discovery		1	
					19
Customer Development	Customer lifetime value		5		
		Classification		1	
		Clustering		2	
		Forecasting		1	
		Regression		1	
	Market basket analysis		3		
		Association		2	
		Sequence		1	
		Discovery			
	Up/cross selling		2		
		Association		1	
		Sequence		1	
		Discovery			
					10
Total			42	42	42

INFERENCE:

Among the four CRM dimensions, customer retention (19 out of 42 banks, 45.23%) is the most common dimension for which data mining is used to support decision making and in that 9 out of 19 banks(47.37%) do one on one marketing in majority for customer retention.

TABLE NO: 6.2

DISTRIBUTION OF BANKS BY DATA MINING TECHNIQUES

Source: Secondary Data

Data mining techniques	Number of Bank	Percentage
Neural network	15	18.52%
Decision tree	11	13.58%
Association rules	9	11.11%
Regression	8	9.88%
Genetic algorithm	3	3.70%
Markov chain	3	3.70%
Survival analysis	3	3.70%
K means	2	2.47%
K nearest neighbour	2	2.47%
Bayesian network classifier	1	1.23%
If-then-else	1	1.23%
Set theory	1	1.23%
Support vector machine	1	1.23%
Attribute oriented induction	1	1.23%
Constructive assignment	1	1.23%
Customer map	1	1.23%
Data envelopment analysis	1	1.23%
Data mining by evolutionary learning	1	1.23%
Expectation Max	1	1.23%
Expectation Max	1	1.23%
Mod Farthest first	1	1.23%
Goal oriented sequential pattern	1	1.23%
Latent class model	1	1.23%
Logical analysis of data	1	1.23%
MARFS1/S2	1	1.23%
Mixture transition distribution	1	1.23%
Multi-classifier class combiner	1	1.23%
Multivariate adaptive regression splines	1	1.23%
Online analytical mining	1	1.23%
Outlier detection	1	1.23%
Pattern based cluster	1	1.23%
Rule-based RIPPER	1	1.23%
S-means	1	1.23%
S-means Mod	1	1.23%
Total	**81**	**100.00%**

INFERENCE:

Among 34 data mining techniques which have been applied in CRM, neural networks is the most commonly used technique. It has been applied in 15 (19.52%) out of 42 banks in total. Following are decision tree and association rules which have been applied in 11 (13.58%) and 9 (11.11%) banks respectively.

TABLE NO: 6.3

DISTRIBUTION OF BANKS BY SOFTWARE / TOOLS USED FOR DATA MINING

Software / Tool Used	Number of Bank	Percentage
Sisense	5	8.47%
Looker	1	1.69%
Zoho Analytics	1	1.69%
Ycllowfin	1	1.69%
Wyn Enterprise	1	1.69%
Qualtrics Research Core	1	1.69%
CXAIR Platform	2	3.39%
Oracle Data Mining	9	15.25%
BigR	5	8.47%
RapidMincr	1	1.69%
Microsoft SharcPoint	7	11.86%
IBM Cognos	5	8.47%
KNIME Analytics Platform	2	3.39%
Dundas BI	4	6.78%
Board	1	1.69%
Orange	2	3.39%
SPSS Modeller	4	6.78%
Big Insights	3	5.08%
Hadoop	4	6.78%
Total	**59**	**100.00%**

Source: Secondary Data

INFERENCE:

Among 19 software tools which are being used by 42 banks, Oracle Data Mining and Microsoft SharePoint is being used in 27.19% banks, that is Oracle Data Mining is being used by 9 banks and Microsoft SharePoint is being used by 7 banks.

FINDINGS

Following are the key findings:

a. Customer retention (19 out of 42 banks, 45.23%) is the most common dimension for which data mining is used to support decision making
b. 9 out of 19 banks (47.37%) who choose Customer Retention as their priority do one on one marketing in majority for customer retention.
c. 23.81% banks choose Customer development as their priority to increase the customers lifetime value.
d. Among 34 data mining techniques which have been applied in CRM, neural networks is the most commonly used technique. It has been applied in 15 (19.52%) out of 42 banks in total.
e. After Neural networks, decision tree and association rules are the most commonly adopted techniques by the banks, that is they have been applied in 11 (13.58%) and 9 (11.11%) banks respectively.
f. Most of the banks do not only rely on one technique, they use more than one technique for data mining. Like State Bank of India uses 4 techniques i.eBig R, SPSS Modeller, Big Insights, and Hadoop.
g. Among 19 software tools which are being used by 42 banks, Oracle Data Mining and Microsoft SharePoint is being used in 27.19% banks, that is Oracle Data Mining is being used by 9 banks and Microsoft SharePoint is being used by 7 banks.
h. Some of the tools are being used in different banks, they have been selected according to the CRM policies of the banks.
i. Choice of Tool is influenced by the type of data banks want to mine.

CONCLUSION

Data mining is an efficient tool to extract knowledge from existing databases. In Banking, data mining plays a vital role in handling transaction data and customer profile. A User can make an effective decision by using data mining techniques. Two major areas of banking application are Customer Relationship Management and Fraud detection. Here the main techniques used are clustering and Association methods.
In this research, based on these methods various types of algorithms are discussed. Finally, we conclude that Bank will obtain a massive profit if they implement data mining in their process of data and decisions.

SUGGESTIONS

a. For Customer Identification, Banks should focus more on target customer analysis to get the best result out of that.
b. As most of the banks are targeting customer retention, so in that area, they should also focus more on complaint management because their focus is tilting towards Loyalty Program and One to One Marketing.
c. For the banks, who want to focus on customer development, they should focus more on up/cross selling because that can provide their customers more benefits.
d. Banks should use more data mining techniques to get better results which can help the banks in providing better customized service.
e. Small Banks should learn from big banks how they are using different tools for analysing the mined data and able to give the best customized services to their customers.
f. More budget and time should be allotted for increasing Customer Relationship by the small banks.

LIMITATION OF THE STUDY

In any research conducted there shall be some limitations associated with it. Hence, for the proper understanding of the project it is inevitable to specify the limitations which are mentioned below:

a. Most of the data mining techniques used by banks are not in public domain.
b. The study is based on quality and originality of secondary data taken through the different portals is considered as another limitation of study.
c. Non-English publications were excluded in this study.
d. There can be more techniques which could have not been considered.

Bibliography

- Alireza Poorebrahimi, Mohammad Ali Afshar Kazemi, Negar Estemda, (1999), Principles of customer relationship management. Student Edn. Mason, Ohio: Thomson/South-Western
- Kurniati, (2000), Data Mining Techniques: for marketing, sales, and customer relationship management. 2nd ed. Indianapolis, Ind: Wiley Pub.
- Chris Rygielski, Jyun Cheng Wang, David C Yen (2002), Data mining for effective risk analysis in a bank intelligence scenario. Proceedings of the 23rd International Conference on Data Engineering Workshop, Apr. 17-20, IEEE Xplore Press, Istanbul, pp: 904-911. DOI: 10.1109/ICDEW.2007.4401083
- Amir M. Hormozi and Stacy Giles, (2004), Analysis of credit card fraud detection methods. International Journal of Recent Trends in Engineering, 2(3), pp.126–128.
- Savitha S. Kadiyal, Alok Srivastava, (2005) From Data Mining to Knowledge Discovery in Databases, AI magazine, 17(3), p.37.
- D. Sunil Anitha, Dr. S. Arumuga Perumal, (2008), The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, 39(11), pp.27– 34.
- Kazi Imran Moin, Dr. Qazi Baseer Ahmed, (2012), CRM at the Speed of Light: Capturing and Keeping Customers in Internet Real Time. Mcgraw-Hill/Osborne, USA, 1st edition
- KeyvanVahidy Rodpysh, amirAghai and Meysam Majdi, (2012), Data Mining: Concepts and Techniques. 3rd Edition, Burlington, MA: Elsevier.
- Dr.RushdyWady, (2013), Customer Management as the origin of Collaborative Customer Relationship Management. Springer, London, UK