DATA MINING : A SYSTEMATIC APPROACH FOR CRM AND OTHER BUSINESS APPLICATIONS

Prof. B. H. Barhate
M.C.S., M.B.A
HOD,Computer Science,
P.O.Nahata College, Bhusawal
I/C Director, IMCD,Bhusawal

ABSTRACT

Applications of Data Mining ***Data Mining*** *has been applied in a wide variety of areas Data mining, data exploration and knowledge discovery are all terms that create an image of the demanding and sometimes tedious search to uncover insights that are neither obvious to competitors nor easy for competitors to duplicate Customer relationship management depends on data analysis activities to uncover directions and opportunities and highlight warning indicators for CRM initiatives. CRM uses data mining to understand how to reach out to and communicate with customers. Data analyses can range from simple, intuitive determination of who to contact, when and where to applying complex algorithms in real-time to deliver customized responses to customer-initiated interaction. Let's review two broad categories of data analysis and see how they might be used to prioritize CRM initiatives. Data Warehouse concept is very useful in Data Mining for CRM*

The official definition provided by DAMA or Boucher: "Data Resource Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise

Data mining is the process of extracting patterns from data. As more data are gathered, with the amount of data doubling every three years data mining is becoming an increasingly important tool to transform these data into information. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery.

While data mining can be used to uncover patterns in data samples, it is important to be aware that the use of non-representative samples of data may produce results that are not indicative of the domain. Similarly, data mining will not find patterns that may be present in the domain, if those patterns are not present in the sample being "mined". There is a tendency for insufficiently knowledgeable "consumers" of the results to attribute "magical abilities" to data mining, treating the technique as a sort of all-seeing crystal ball. Like any other tool, it only functions in conjunction with the appropriate raw material: in this case, indicative and representative data that the user must first collect. Further, the discovery of a particular pattern in a particular set of

data does not necessarily mean that pattern is representative of the whole population from which that data was drawn. Hence, an important part of the process is the verification and validation of patterns on other samples of data.

The term data mining has also been used in a related but negative sense, to mean the deliberate searching for apparent but not necessarily representative patterns in large numbers of data. To avoid confusion with the other sense, the terms *data dredging* and *data snooping* are often used. Note, however, that dredging and snooping can be (and sometimes are) used as exploratory tools when developing and clarifying hypotheses.

Concepts of Data Mining in CRM :

Data mining, *the extraction of hidden predictive information from large databases,* is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"
The way in which companies interact with their customers has changed dramatically over the past few years. A customer's continuing business is no longer guaranteed. As a result, companies have found that they need to understand their customers better, and to quickly respond to their wants and needs. In addition, the time frame in which these responses need to be made has been shrinking. It is no longer possible to wait until the signs of customer dissatisfaction are obvious before action must be taken. To succeed, companies must be proactive and anticipate what a customer desires.

Customers and prospective customers want to interact on their terms, meaning that you need to look at multiple criteria when evaluating how to proceed. You will need to automate:

- The Right Offer
- To the Right Person
- At the Right Time
- Through the Right Channel

Customer relationship management (CRM) is a process that manages the interactions between a company and its customers. The primary users of CRM software applications are database marketers who are looking to automate the process of interacting with customers.

To be successful, database marketers must first identify market segments containing customers or prospects with high-profit potential. They then build and execute campaigns that favorably impact the behavior of these individuals.

The first task, identifying market segments, requires significant data about prospective customers and their buying behaviors. In theory, the more data the better. In practice, however, massive data stores often impede marketers, who struggle to sift through the minutiae to find the nuggets of valuable information.
In the past, the link between data mining and campaign management software was mostly manual. In the worst cases, it involved "sneaker net," creating a physical file on tape or disk, which someone then carried to another computer and loaded into the marketing database.

Data mining in customer relationship management applications can contribute significantly to the bottom line. Rather than randomly contacting a prospect or customer through a call center or sending mail, a company can concentrate its efforts on prospects that are predicted to have a high likelihood of responding to an offer. More sophisticated methods may be used to optimise resources across campaigns so that one may predict which channel and which offer an individual is most likely to respond to — across all potential offers. Additionally, sophisticated applications could be used to automate the mailing. Once the results from data mining (potential prospect/customer and channel/offer) are determined, this "sophisticated application" can either automatically send an e-mail or regular mail. Finally, in cases where many people will take an action without an offer, uplift modeling can be used to determine which people will have the greatest increase in responding if given an offer. Data clustering can also be used to automatically discover the segments or groups within a customer data set.

Businesses employing data mining may see a return on investment, but also they recognise that the number of predictive models can quickly become very large. Rather than one model to predict which customers will churn, a business could build a separate model for each region and customer type. Then instead of sending an offer to all people that are likely to churn, it may only want to send offers to customers that will likely take to offer. And finally, it may also want to determine which customers are going to be profitable over a window of time and only send the offers to those that are likely to be profitable. In order to maintain this quantity of models, they need to manage model versions and move to *automated data mining.*

CRM, or Customer Relationship Management, is a business strategy of a company to identify and manage the customer by reducing company's costs and increasing profitability. CRM redesign the basic functional activities of a company that in turn demands re-engineering of work processes. To have successful CRM, it is necessary to bring together all the information from all possible sources (within / outside organization) about the customer and process in such a manner that the organization gets a real about the customer behaviour. This will help the management, sales people, marketing and people providing service so that they can react to the customer accordingly and keep pace with the competition. It is practically impossible to consider these processes without addressing technology. CRM applications can enable effective Customer Relationship Management, provided that an organization has the right

business strategy. Three factors are responsible for the successful implementation of CRM: People, Process and Technology.

People: In an organization starting from top layer of employees to bottom layer employees should support CRM and promote the same.

Process: Business Process should be re-engineered.

Technology: Right technology should be chosen to drive these new processes, provide real time customer analysis to the employees.

If the business strategy is clearly defined, it is the technology that can drive the business and customer services more efficiently with the full co-operation of people.

Data Mining plays an important role in the implementation of Analytical CRM. Analytical CRM comprises all programming that analyzes data about an enterprise's customers and presents it so that better and quicker business decisions can be made by employing online analytical processing (OLAP) and Data Mining tool.

Before getting into details of Data Mining let's discuss about the building up data warehouse.

DATA WAREHOUSE

Date warehousing means a central repository of all critical data, which help managers to take decisions, based on authentic information. Building a data warehouse is not an easy task. The type of data to be kept in a data warehouse is a pivotal issue for the organisation; since this exercise involves a lengthy and tedious process of consolidating all back end data from different databases.

A Data Warehouse supports business analysis and decision-making process by creating an integrated database of consistent, subject-oriented and historical information. A Data warehouse is a database, populated from the existing data source system and used for reporting, analysis & business intelligence purposes. It integrates data from any kind of data structure including external data into one consolidated database. By transforming and integrating data into meaningful information, a data warehouse helps business analysts in taking business decisions. Business trends can be forecast by analyzing the historical data.

W. H. Inmon, the 'father' of data warehousing, or other authorities in the industry, the definition of a data warehouse is a subject-oriented, integrated, time variant, non-volatile collection of data in support of management decisions.

- **Subject-Oriented** means that all relevant data organized and summarised about a topic such as Banking, Insurance, Marketing, etc. For each of these topics the data warehouse has a specific subject for analysis such as customer, product, etc.

- **Integrated** refers data stored in a standarised format by integrating all data structures including external data

- **Time Variant** represents long-term data

- **Non-Volatile** means data doesn't change once the data is in the warehouse.

Data Warehouse Benefits

Implementing a Data Warehouse provides significant benefits -- many tangible, some intangible. They are:

- Decreased costs of product introduction with improved selection of target markets
- Better business enterprise intelligence
- Enhanced customer service
- Business Process and Information System Re-engineering

Structure of a Data Warehouse

Data Warehouses have a distinct structure. There are different levels of summarization and detail that describe the Data Warehouse. The different components of the Data Warehouse are:

- **Current Detail Data**: The heart of a data warehouse is its current detail data, which is acquired directly from the operational databases, and may be stored as raw data or as aggregations of raw data. The current detail data is organized along subject area represents the entire enterprise, rather than a given application.
- **Older Detail Data**: This represents the aged current detail data, or the history of the subject areas; this data is what makes trend analysis possible.
- **Summarized Data**: The data in a data warehouse is stored in a summarised form so that the data can be easily accessed and analysed without time-consuming manipulation and processing. Decisions can be made more quickly and with confidence that the data is both timely and accurate. The summarised data is classified as Lightly Summarised Data and Highly Summarised Data.

    - **Lightly Summarised Data** is data that is highly filtered data as all the data is not required to be stored in a warehouse
    - **Highly Summarised Data** is compact and easily accessible.

- - **Meta Data**: Meta Data is an important component of the data warehouse that facilitates the user to know what is contained in the warehouse and in fact guides the use through the warehouse. To handle the links between stored data we have to define the relationships that is defined as Metadata whereas data is defined as Facts. *Without metadata, data is likely to be useful only to the persons who collected the data in the first place, as they are the only ones who understand the frame of reference that goes with the data.* So, Metadata is "data about data" - it is not the actual dataset, but answers the "who, what, where, when, why and how" questions about the dataset.

  Meta Data can be classified into two categories:
    - Technical Meta Data
    - Information about data sources
    - Transformation description
    - Warehouse object and data structure definitions
    - Rules to perform data clean up and data enhancement
    - Data mapping
    - Access authorization and history

    - Business Meta Data
    - Subject areas and information object type
    - Internet Homepages
    - Other information to support all data warehousing components

- Data warehouse operational information

Data Mart

A data mart is a repository of data gathered from operational data and other sources that is designed to serve a particular community of knowledge workers. In scope, the data may derive from an enterprise-wide database or data warehouse or be more specialized. The emphasis of a data mart is on meeting the specific demands of a particular group of knowledge users in terms of analysis, content, presentation, and ease-of-use. Users of a data mart can expect to have data presented in terms that are familiar.

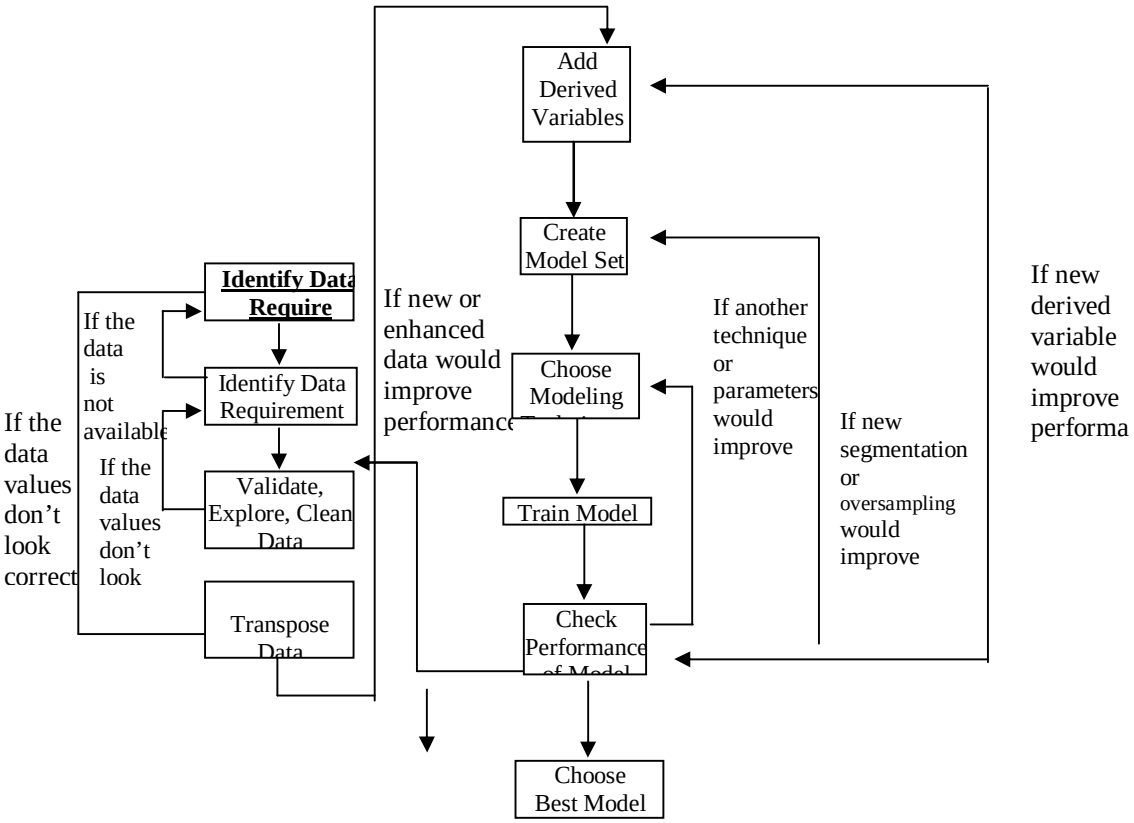
In practice, the terms *data mart* and *data warehouse* each tend to imply the presence of the other in some form. However, most writers using the term seem to agree that the design of a data mart tends to start from an analysis of user needs and that a data warehouse tends to start from an analysis of what data already exists and how it can be collected in such a way that the data can later be used. A data warehouse is a central aggregation of data (which can be distributed physically); a data mart is a data repository that may derive from a data warehouse or not and that emphasizes ease of access and usability for a particular designed purpose. In general, a data warehouse tends to be a strategic but somewhat unfinished concept; a data mart tends to be tactical and aimed at meeting an immediate need.

Building a Data Warehouse

In general, building any data warehouse require the following steps:

- **Data Warehouse Design**: The organisation has to choose appropriate architecture for designing Data Warehouse. They are Star, Snow Flake, and Entity-Relationship. And also to identify the summarised data to be pulled from transactional database to data warehouse.
- **Extracting Transactional Data**: Extraction of data from internal and external sources and apply transformations as per business rules and put in a structured and standarised format, then load into the Data Warehouse.
- **Data Availability**: Data is to be available for analytic application and decision-making application.
- **Data Warehouse Strategy**: The strategy should be such that it should help in achieving business goals and the data warehouse should be flexible enough to confirm to the business need as and when business strategy changes. The strategy should include how long data can be archived?
- **Data Quality**: The most important aspect of Data warehouse is data quality and maintenance. The data pumped into data warehouse should match with the business rule and also measure the data quality. If the qualitative data is not achieved, it is necessary to change the data transformation process to ensure that good quality of data is pushed into data warehouse.

The diagram below illustrates the basic steps in Data mining.



- Identify and obtain data
- Validate, Explore, Clean data
- Transform data to right level of Granularity
- Add derived variables
- Choose Modeling Techniques
- Train the Model
- Check the Model performance
- Choose the best model

Broadly, the types of data used in Data mining applications is given below:

- Demographic data such as age, income, profession etc.
- Transaction data: Specific to application
- Data shared within an Industry: Credit reports, catalogues etc.
- Data shared from Business partners

Data collected for data mining can be grouped under THREE major categories

- Categorical: Set of values, which a field can take. There is no ordering or hierarchy in data. Examples: Profession, Qualification, City, Pin code etc.
- Rank: Similar to categorical data but have a natural ordering. Examples are income range like low, middle, high, Age: Young, middle age, old, very old etc.

The Foundations of Data Mining

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Automated prediction of trends and behaviors. Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. Automated discovery of previously unknown patterns. Data mining tools sweep through databases and identify previously hidden patterns in one step.

Data mining techniques can yield the benefits of automation on existing software and hardware platforms, and can be implemented on new systems as existing platforms are upgraded and new products developed. When data mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means that users can automatically experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions.

Knowledge Discovery in Databases (KDD) is the name coined by Gregory Piatetsky-Shapiro in 1989 to describe the process of finding interesting, interpreted, useful and novel data. There are many nuances to this process, but roughly the steps are to preprocess raw data, mine the data, and interpret the results

The most commonly used techniques in data mining are:

Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

Rule induction: The extraction of useful if-then rules from data based on statistical significance.

Data mining has clearly moved into the mainstream. This approach to discovering previously unknown patterns or connections in data was developed in academia and first employed by government research labs. Now it plays a central role in helping companies in virtually every industry improve daily business decision making. And one of the most effective uses of data mining is to support improvements in customer relationship management (CRM). More cost-effective customer acquisition and retention, better targeting of marketing campaigns, improved cross-selling and up-selling to increase customer value—these are just some of the customer issues that leading companies address through data mining. Data mining is a key component of predictive analytics. Through predictive analytics, your company can use insights gained by analyzing data to direct, optimize, and automate customer-focused interactions, anywhere in your organization.

If your company is relatively new to data mining, you can benefit from the experience of companies that have used it successfully. They have discovered a number of tips and tricks to achieving the maximum return on their investment (ROI). What follows is a distillation of this experience: "The Top 10 Secrets to Using Data Mining to Succeed at CRM."

1. **Planning is the key to a successful data mining project**
2. **Set specific goals for your data mining project**
3. **Recruit a broad-based project team**
4. **Line up the right data**
5. **Secure IT buy-in**
6. **Select the right data mining solution**
7. **Consider mining other types of data to increase the return on your data mining investment**
8. **Expand the scope of data mining to achieve even greater results**
9. **Consider all available deployment options**
10. **Increase collaboration and efficiency through model management**

How Data Mining Works

How exactly is data mining able to tell you important things that you didn't know or what is going to happen next? The technique that is used to perform these feats in data mining is called modeling. Modeling is simply the act of building a model in one situation where you know the answer and then applying it to another situation that you don't.

This act of model building is thus something that people have been doing for a long time, certainly before the advent of computers or data mining technology. What happens on computers, however, is not much different than the way people build models. Computers are loaded up with lots of information about a variety of situations where an answer is known and then the data mining software on the computer must run through that data and distill the characteristics of the data that should go into the model. Once the model is built it can then be used in similar situations where you don't know the answer

This model could then be applied to the prospect data to try to tell something about the proprietary information that this telecommunications company does not currently have access to. With this model in hand new customers can be selectively targeted.

Profitable Applications

A wide range of companies have deployed successful applications of data mining. While early adopters of this technology have tended to be in information-intensive industries such as financial services and direct mail marketing, the technology is applicable to any company looking to leverage a large data warehouse to better manage their customer relationships. Two critical factors for success with data mining are: a large, well-integrated data warehouse and a well-defined understanding of the business process within which data mining is to be applied (such as customer prospecting, retention, campaign management, and so on).

- A large consumer package goods company can apply data mining to improve its sales process to retailers. Data from consumer panels, shipments, and competitor activity can be applied to understand the reasons for brand and store switching. Through this analysis, the manufacturer can select promotional strategies that best reach their target customer segments.
- In Marketing it helps to collect historical Data from all types of customers.
- Easy way to gather statistical analysis using Data Warehouse.

These Application leverage the knowledge about customers implicit in a data warehouse to reduce costs and improve the value of customer relationships. These organizations can now focus their efforts on the most important (profitable) customers and prospects, and design targeted marketing strategies to best reach them.

References

1. Database Management System Concept by Henry Korth
2. E-commerce papers – proceeding, M.J.College Jalgaon
3. ICDL2004 , International Conference on Digital Library New Delhi
   Papers on e-commerce
4. http://en.wikipedia.org/wiki/Data_mining
5. E-commerce by CVS Murty
6. Ref . Paper of B.H.Barhate , Title CRM at IIM Indore Conference
7. Ref. Paper of B.H.Barhate , Title Knowledge Management at Mumbai
8. http://i.zdnet.com/whitepapers/SPSS_The_Top_10_Secrets_to_Using_Data_Mining_to_Succeed_at_CRM_whitepaper.pdf
9. http://www.niapune.com/pdfs/Bimaquest/Volume-3_Issue%202/das.htm