DV Digital Valley

Editor : Sohan Khemka | Digital Valley - Monthly

Editor's Point of View…

Why Data Mining?

It is important to note that time is spent in getting valid information from the data. Therefore if you are after making your business grow rapidly, there is a need to make accurate and quick decisions that can take advantage of grabbing the available opportunities in a timely manner.

Data mining is a rapidly growing industry in this technology trend world. Everyone now-a-days required their data to be used in an appropriate manner and up to right approach in order to obtain useful and accurate information.

Governments, private companies, large organizations and all businesses are after a large volume of data collection for the purposes of business and research development. The data collected can be stored for future use. Storage of information is quite important whenever it is required. It is important to note that it may take a long time for finding and searching for information from websites, databases and other internet sources.

Why ERP?

Every business has its own objectives, processes and requirements. Above all, today's businesses need technologies with complete functions which can bridge the gap between business processes and people. To run a large organization with multiple departments and teams successfully, an ERP system gives a helping hand by synchronizing all information and communication within the organization. ERP is a combination of software and company's activities performed to manage operations. With ERP software, the entire project value chain is aligned and critical processes are streamlined effectively.

So, why and when do you need an ERP system? When your business starts dropping the ball, lacking proper communications, or you observe you are giving more time in paperwork than in actually running the business, you need to have an ERP system. You need to set up various workflows and methods (automated & manual) which may or may not be documented formally.

ERP systems allow companies to achieve actual business process automation, managing important tasks on a daily basis across an organization and freeing up staff to focus their efforts on more critical initiatives which require personal attention.

Regards,
Sohan Khemka

CONTENTS

How Data Mining Can Help Businesses To Get A Competitive Edge?

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 and extract useful information from it.

Extraction of information is not the only process we need to perform; data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these 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.

What Is DATA MINING?

Data Mining is defined as extracting information from huge sets of data. In other words, we can say that data mining is the procedure of mining knowledge from data.

Data Mining Usage in Various Domains:

Data mining is highly useful in the following domains
- Market Analysis and Management
- Corporate Analysis & Risk Management
- Fraud Detection

Apart from these, data mining can also be used in the areas of production control, customer retention, science exploration, sports, astrology, and Internet Web Surf-Aid

Data Mining In Market Analysis and Management

Listed below are the various fields of market where data mining is used:

- **Customer Profiling** – Data mining helps determine what kind of people buy what kind of products.
- **Identifying Customer Requirements** – Data mining helps in identifying the best products for different customers. It uses prediction to find the factors that may attract new customers.
- **Cross Market Analysis** – Data mining performs Association/correlations between product sales.
- **Target Marketing** – Data mining helps to find clusters of model customers who share the same characteristics such as interests, spending habits, income, etc.
- **Determining Customer purchasing pattern** – Data mining helps in determining customer purchasing pattern.
- **Providing Summary Information** – Data mining provides us various multidimensional summary reports.

Data Mining In Corporate Analysis and Risk Management

Data mining is used in the following fields of the Corporate Sectors.
- **Finance Planning and Asset Evaluation** – It involves cash flow analysis and prediction, contingent claim analysis to evaluate assets.
- **Resource Planning** – It involves summarizing and comparing the resources and spending.
- **Competition** – It involves monitoring competitors and market directions.

Data Mining In Fraud Detection

Data mining is also used in the fields of credit card services and telecommunication to detect frauds. In fraud telephone calls, it helps to find the destination of the call, duration of the call, time of the day or week, etc. It also analyses the patterns that deviate from expected norms.

Types of Data

Data mining can be performed on following types of data

- Relational databases
- Data warehouses
- Advanced DB and information repositories
- Object-oriented and object-relational databases
- Transactional and Spatial databases
- Heterogeneous and legacy databases
- Multimedia and streaming database
- Text databases
- Text mining and Web mining

Data Mining Techniques

1. Classification:

This analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes.

2. Clustering:

Clustering analysis is a data mining technique to identify data that are like each other. This process helps to understand the differences and similarities between the data.

3. Regression:

Regression analysis is the data mining method of identifying and analysing the relationship between variables. It is used to identify the likelihood of a specific variable, given the presence of other variables.

4. Association Rules:

This data mining technique helps to find the association between two or more Items. It discovers a hidden pattern in the data set.

5. Outer detection:

This type of data mining technique refers to observation of data items in the dataset which do not match an expected pattern or expected behaviour. This technique can be used in a variety of domains, such as intrusion, detection, fraud or fault detection, etc. Outer detection is also called Outlier Analysis or Outlier mining.

6. Sequential Patterns:

This data mining technique helps to discover or identify similar patterns or trends in transaction data for certain period.

7. Prediction:

Prediction has used a combination of the other data mining techniques like trends, sequential patterns, clustering, classification, etc. It analyses past events or instances in a right sequence for predicting a future event.

Challenges of Implementation of Data mine:

- Skilled Experts are needed to formulate the data mining queries.
- Overfitting: Due to small size training database, a model may not fit future states.
- Data mining needs large databases which sometimes are difficult to manage
- Business practices may need to be modified to determine to use the information uncovered.
- If the data set is not diverse, data mining results may not be accurate.
- Integration information needed from heterogeneous databases and global information systems could be complex

Data mining Examples

Consider a marketing head of telecom service provides who wants to increase revenues of long distance services. For high ROI on his sales and marketing efforts customer profiling is important. He has a vast data pool of customer information like age, gender, income, credit history, etc. But its impossible to determine characteristics of people who prefer long distance calls with manual analysis. Using data mining techniques, he may uncover patterns between high long distance call users and their characteristics.

A bank wants to search new ways to increase revenues from its credit card operations. They want to check whether usage would double if fees were halved.

Bank has multiple years of record on average credit card balances, payment amounts, credit limit usage, and other key parameters. They create a model to check the impact of the proposed new business policy. The data results show that cutting fees in half for a targeted customer base could increase revenues by $10 million.

Benefits of Data Mining:

- Data mining technique helps companies to get knowledge-based information.
- Data mining helps organizations to make the profitable adjustments in operation and production.
- The data mining is a cost-effective and efficient solution compared to other statistical data applications.
- Data mining helps with the decision-making process.
- Facilitates automated prediction of trends and behaviours as well as automated discovery of hidden patterns.
- It can be implemented in new systems as well as existing platforms
- It is the speedy process which makes it easy for the users to analyse huge amount of data in less time.
- In finance and banking, data mining is used to create accurate risk models for loans and mortgages. They are also very helpful when detecting fraudulent transactions.

Disadvantages of Data Mining

- There are chances of companies may sell useful information of their customers to other companies for money. For example, American Express has sold credit card purchases of their customers to the other companies.
- Many data mining analytics software is difficult to operate and requires advance training to work on.
- Different data mining tools work in different manners due to different algorithms employed in their design. Therefore, the selection of correct data mining tool is a very difficult task.
- The data mining techniques are not accurate, and so it can cause serious consequences in certain conditions.

Data Mining Implementation Process

Let's study the Data Mining implementation process in detail

Business Understanding:

In this phase, business and data-mining goals are established.

- First, you need to understand business and client objectives. You need to define what your client wants (which many times even they do not know themselves)
- Take stock of the current data mining scenario. Factor in resources, assumption, constraints, and other significant factors into your assessment.
- Using business objectives and current scenario, define your data mining goals.
- A good data mining plan is very detailed and should be developed to accomplish both business and data mining goals.

Data Understanding

In this phase, sanity check on data is performed to check whether it's appropriate for the data mining goals.

- First, data is collected from multiple data sources available in the organization.
- These data sources may include multiple databases, flat filer or data cubes. There are issues like object matching and schema integration which can arise during Data Integration process. It is a quite complex and tricky process as data from various sources unlikely to match easily. For example, table A contains an entity named cust_no whereas another table B contains an entity named cust-id.
- Therefore, it is quite difficult to ensure that both of these given objects refer to the same value or not. Here, Metadata should be used to reduce errors in the data integration process.
- Next, the step is to search for properties of acquired data. A good way to explore the data is to answer the data mining questions (decided in business phase) using the query, reporting, and visualization tools.
- Based on the results of query, the data quality should be ascertained. Missing data if any should be acquired.

Data preparation

In this phase, data is made production ready.

The data preparation process consumes about 90% of the time of the project.

The data from different sources should be selected, cleaned, transformed, formatted, anonymised, and constructed (if required).

Data cleaning is a process to "clean" the data by smoothing noisy data and filling in missing values.

For example, for a customer demographics profile, age data is missing. The data is incomplete and should be filled. In some cases, there could be data outliers. For instance, age has a value 300. Data could be inconsistent. For instance, name of the customer is different in different tables.

Data transformation operations change the data to make it useful in data mining. Following transformation can be applied:

Data transformation:

Data transformation operations would contribute toward the success of the mining process.

Smoothing: It helps to remove noise from the data.

Aggregation: Summary or aggregation operations are applied to the data. I.e., the weekly sales data is aggregated to calculate the monthly and yearly total.

Generalization: In this step, Low-level data is replaced by higher-level concepts with the help of concept hierarchies. For example, the city is replaced by the county.

Normalization: Normalization performed when the attribute data are scaled up o scaled down. Example: Data should fall in the range -2.0 to 2.0 post-normalization.

Attribute construction: these attributes are constructed and included the given set of attributes helpful for data mining.

The result of this process is a final data set that can be used in modelling.

Modelling

In this phase, mathematical models are used to determine data patterns.

- Based on the business objectives, suitable modelling techniques should be selected for the prepared dataset.
- Create a scenario to test check the quality and validity of the model.
- Run the model on the prepared dataset.
- Results should be assessed by all stakeholders to make sure that model can meet data mining objectives.

Evaluation

In this phase, patterns identified are evaluated against the business objectives.

- Results generated by the data mining model should be evaluated against the business objectives.
- Gaining business understanding is an iterative process. In fact, while understanding, new business requirements may be raised because of data mining.
- A go or no-go decision is taken to move the model in the deployment phase.

Deployment

In the deployment phase, you ship your data mining discoveries to everyday business operations.

- The knowledge or information discovered during data mining process should be made easy to understand for non-technical stakeholders.

- A detailed deployment plan, for shipping, maintenance, and monitoring of data mining discoveries is created.

- A final project report is created with lessons learned and key experiences during the project. This helps to improve the organization's business policy.

Following a systematic approach for data mining implementation can greatly reduce the risks of project failure. Moreover, it can help business and technical folks to determine the need of data analytics and the best tools and techniques to choose from.

Where is Data Mining used?

Applications	Usage
Communications	Data mining techniques are used in communication sector to predict customer behaviour to offer highly targeted and relevant campaigns.
Insurance	Data mining helps insurance companies to price their products profitable and promote new offers to their new or existing customers.
Education	Data mining benefits educators to access student data, predict achievement levels and find students or groups of students which need extra attention. For example, students who are weak in maths subject.
Manufacturing	With the help of Data Mining Manufacturers can predict wear and tear of production assets. They can anticipate maintenance which helps them reduce them to minimize downtime.
Banking	Data mining helps finance sector to get a view of market risks and manage regulatory compliance. It helps banks to identify probable defaulters to decide whether to issue credit cards, loans, etc.
Retail	Data Mining techniques help retail malls and grocery stores identify and arrange most sellable items in the most attentive positions. It helps store owners to come up with the offer which encourages customers to increase their spending.
Service Providers	Service providers like mobile phone and utility industries use Data Mining to predict the reasons when a customer leaves their company. They analyse billing details, customer service interactions, complaints made to the company to assign each customer a probability score and offers incentives.
E-Commerce	E-commerce websites use Data Mining to offer cross-sells and up-sells through their websites. One of the most famous names is Amazon, who use Data mining techniques to get more customers into their ecommerce store.
Super Markets	Data Mining allows supermarkets develop rules to predict if their shoppers were likely to be expecting. By evaluating their buying pattern, they could find woman customers who are most likely pregnant. They can start targeting products like baby powder, baby shop and diapers and so on.
Crime Investigation	Data Mining helps crime investigation agencies to deploy police workforce (where is a crime most likely to happen and when?), who to search at a border crossing etc.
Bioinformatics	Data Mining helps to mine biological data from massive datasets gathered in biology and medicine.

Enterprise Resource Planning (ERP)

Introduction

In any industry, some of the demands managers face is to be cost effective. In addition to that, they are also faced with challenges such as to analyse costs and profits on a product or consumer basis, to be flexible to face ever altering business requirements, and to be informed of management decision making processes and changes in ways of doing business.

However, some of the challenges holding managers back include the difficulty in attaining accurate information, lack of applications that mimic existing business practices and bad interfaces. When some challengers are holding a manager back, that is where Enterprise Resource Planning (ERP) comes into play.

Over the years business applications have evolved from Management Information Systems with no decision support to Corporate Information Systems, which offer some decision support to Enterprise Resource Planning. Enterprise Resource Planning is a software solution that tackles the needs of an organization, taking into account the process view to meet an organization's goals while incorporating all the functions of an organization.

Its purpose is to make easy the information flow between all business functions within the boundaries of the organization and manage the organization's connections with its outside stakeholders.

The Driving Force behind ERP

There are two main driving forces behind Enterprise Resource Planning for a business organization.

In a business sense, Enterprise Resource Planning ensures customer satisfaction, as it leads to business development that is development of new areas, new products and new services.

Also, it allows businesses to face competition for implementing Enterprise Resource Planning, and it ensures efficient processes that push the company into top gear.

Enterprise Resource Planning is necessary for management in today's business world.

Advantages of ERP System

- With Enterprise Resource Planning (ERP) software, accurate forecasting can be done. When accurate forecasting inventory levels are kept at maximum efficiency, this allows for the organization to be profitable.
- Integration of the various departments ensures communication, productivity and efficiency.
- Adopting ERP software eradicates the problem of coordinating changes between many systems.
- ERP software provides a top-down view of an organization, so information is available to make decisions at anytime, anywhere.

Disadvantages of ERP System

- Adopting ERP systems can be expensive.
- The lack of boundaries created by ERP software in a company can cause problems of who takes the blame, lines of responsibility and employee morale.

Real Time Example of ERP System

eduERP® - eduERP® is a full-fledged ERP specifically designed for educational institutions and governing bodies. It addresses all requirements of an educational institution by providing the necessary infrastructure for eLearning and eGovernance.

Title Verification Code: GUJENG16187 Issued by Registrar of the Newspaper for India, New Delhi
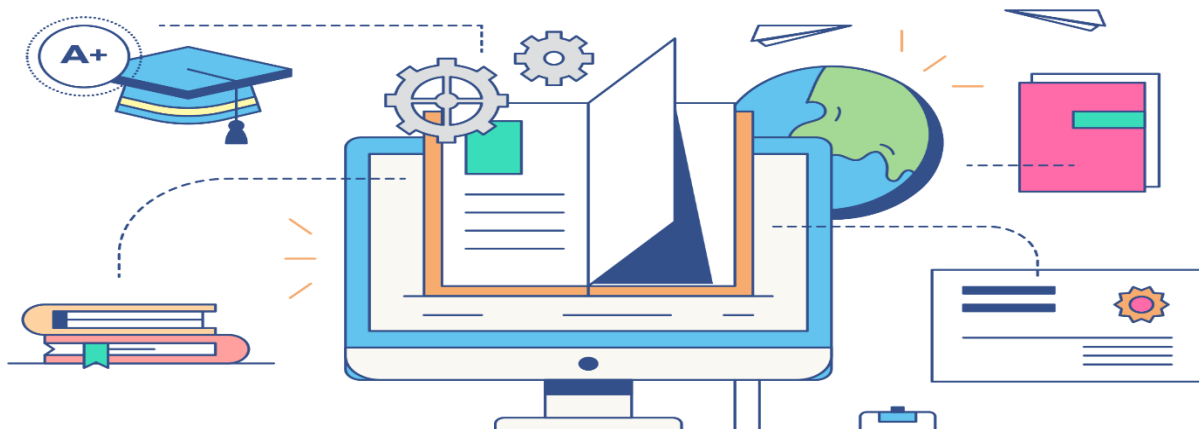
Editor and Publisher are not responsible for the View / opinions expressed in articles. They are View / opinions of author/s.