



Param Pujya Dr. Babasaheb Ambedkar Smarak Samiti's

**Dr. Ambedkar Institute of Management Studies & Research**

Deeksha Bhoomi, Nagpur - 440010 (Maharashtra State) INDIA

NAAC Accredited with 'A' Grade

Tel: +91 712 6521204, 6521203, 6501379

Email: info@daimsr.in

# Data Mining & Data Warehousing

## Unit IV

For Internal Circulation and  
Academic Purpose Only

# Programme Educational Objectives

*Our program will create graduates who:*

- 1. Will be recognized as a creative and an enterprising team leader.*
- 2. Will be a flexible, adaptable and an ethical individual.*
- 3. Will have a holistic approach to problem solving in the dynamic business environment.*

# Business Communication & Information Systems

## Course Outcomes

- CO1-Given the circumstances, student manager will be able to ascertain the barriers to communication and also propose measures to overcome these barriers.
- CO2 In a given situation, student manager will be able to identify essentials parameters of effective communication and will also be able to justify the same.
- CO3 For given situation student manager should be able to draft Business letter for an organization.
- CO4 Given the circumstances, student manager will be able to draft E-mail to concerned authority/person.
- CO5 Given the circumstances, student manager will be able to gather data and make an informed decision based on it.
- CO6 Student manager will be able to identify & explain modern trends in information system.

# Introduction

- What is Data Mining?
- Data Mining is the process of collecting large amounts of raw data and transforming that data into useful information.
- Data Warehousing?
- A Data Warehouse is a computerized collection of mined data.

# What is Data Mining?

- Data mining is the practice of searching through large amounts of computerized data to find useful patterns or trends (American Heritage Dictionary, 2008).

# Data Mining Applications

- Banking
  - Detect Fraudulent Activity
- Insurance
  - Risk Assessment
- Medicine/Healthcare
  - Enhance Research
- Retail
  - Track consumer buying trends

# Data Mining Software Applications

- -SPSS/ SPSS Clementine is a data mining toolkit with a wide variety of mining methods to choose from. This package is designed to be used by the normal person with the desire to do their own data mining projects
- -Previously known as yale, RapidMiner is a data mining suite which makes a wide range of techniques available. . Its easy to use interface makes it accessible for general use, while its flexibility and extensibility make it suitable for academic use and adds a number of useful visualization methods.
- -‘SAS Enterprise Miner,’ a tool provided in the package suite, has an easy to use data-flow GUI. A wide variety of data mining techniques are supported including: Decision Trees; Neural Networks; Market Basket Analysis; Clustering; Web Log Analysis; Regression; and Rule Induction.



# CRISP-DM

## Cross-Industry Standard Process for Data Mining

- Understanding the business
- Understanding the data
- Data preparation
- Modeling
- Evaluation
- Deployment





# Data Warehousing

## *Advantages*

- Access to information
- Data Inconsistency
- Decrease Computing Cost
- Productivity Increase
- Increase company profits

# Data Warehousing

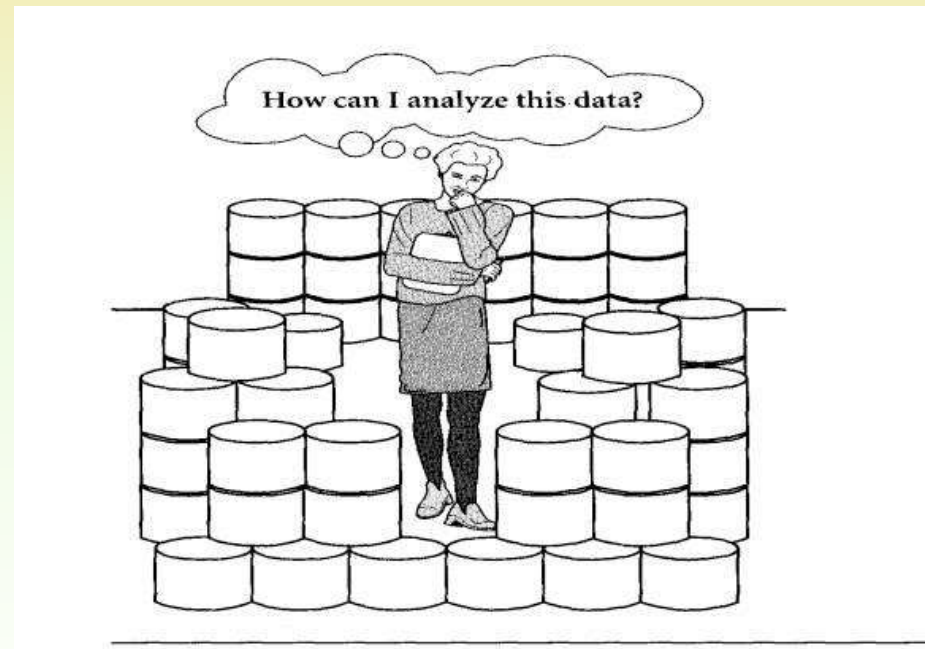
## *Disadvantages*

- Data must be cleaned, loaded, and extracted
  - 80% of the overall process
- User Variability
  - Proper Training
- Difficult to Maintain
  - Incongruence among systems

# Data Mining

## *Advantages*

- Improves Customer Satisfaction/service
- Saves Time and Money
- Increases Sales Effectiveness
- Increases profitability



# Data Mining

## *Disadvantages*

- Require skilled technical users to interpret and analyze data from warehouse
- Validity of the patterns
  - Related to real world circumstances
- Unable to Identify Casual Relationships
- Reserved for the few instead of the many

# Current Issues

## Data Quality

- Duplicated records
- Lack of Data Standards
- Human Error

## Inoperability

- Lack of communications among existing systems

## Mission Creep

# Trends & Current Issues

## •4 Major Trends

- Data – growing amount collected to be sifted
- Hardware – growing performance & storage
- Scientific Computing – theory, experiment, simulation
- Business – Meet higher standard in order to foresee risks, opportunities, and benefits for the company

- Growing quickly due to renewal with new methodology frequently discovered
- Applications of uses & methodology to medical, marketing, operations, & others
- The government is closely reviewing the uses of data mining, due to the possibilities both good and bad
- Counterterrorism data mining has been done, but in some instances has been deemed a violation of privacy



# Future Research Possibilities

- Government's uses for data mining.
  - National Security
  - Terrorism Detection
- Identity theft through data mining.

# Conclusion/Analysis

- Data mining is the extraction of information that can predict future trends & behaviors
- Requires a large amount of data to be collected, and then stored in data warehouse
- Possible violation of privacy in some circumstances
- Government is getting involved with regulation, despite the counterterrorism program being a possible violation

# Reference

- Goyal D. P. Management Information Systems, Managerial Perspectives, Macmillan India Ltd. Second Edition 2008
- Oka, M. M. Management information System 14th Edition, Everest Publication.
- Sadagopan. S. Management information System, PHI publication