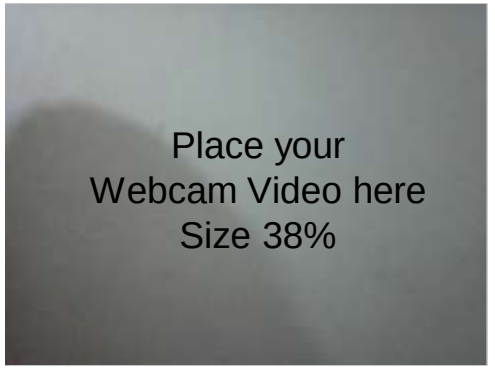

# Data mining

Aiswarya V
Department of Computer Science
Amrita Vishwa Vidyapeetham

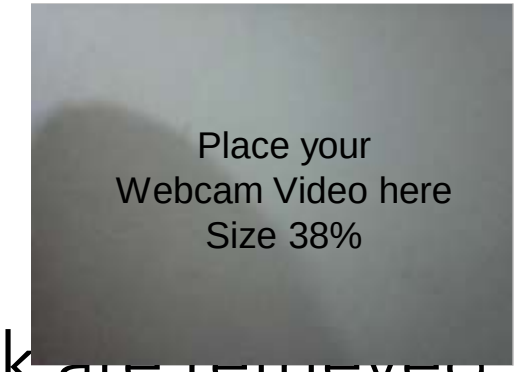
# KDD




- Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data, or KDD, while others view data mining as merely an essential step in the process of knowledge discovery.
- The knowledge discovery process is an iterative sequence of the following steps:

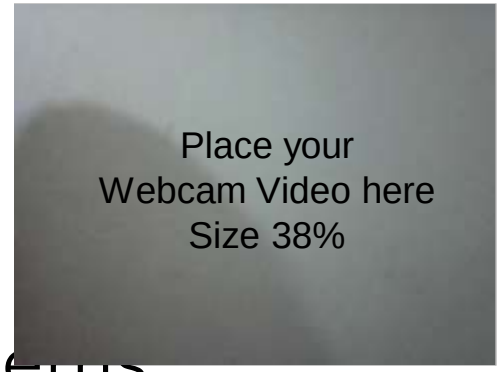
1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)

# <Content Slide>




3. Data selection (where data relevant to the analysis task are retrieved from
the database)
4. Data transformation (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)
5. Data mining (an essential process where intelligent methods are applied to extract data patterns)

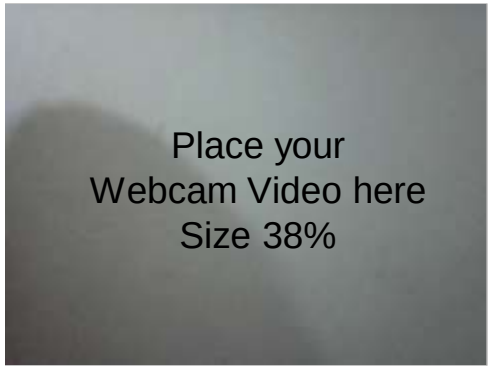
# <Content Slide>




6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interestingness measures)
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users)

Steps 1 through 4 are different forms of data preprocessing, where data are prepared for mining

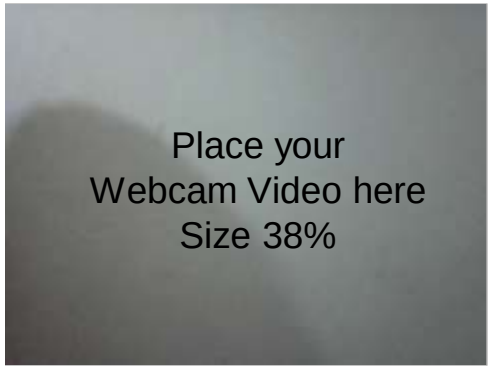
# <Content Slide>




- However, in industry, in media, and in the research milieu, the term data mining is often used to refer to the entire knowledge discovery process (perhaps because the term is shorter than knowledge discovery from data).

- Therefore, we adopt a broad view of data mining functionality: Data mining is the process of discovering interesting patterns and knowledge from large amounts of data.

# Multi-Dimensional View of Data Mining




**Data to be mined**

Database data (extended-relational, object-oriented, heterogeneous, legacy), data warehouse, transactional data, stream, spatiotemporal, time-series, sequence, text and web, multi-media, graphs & social and information networks

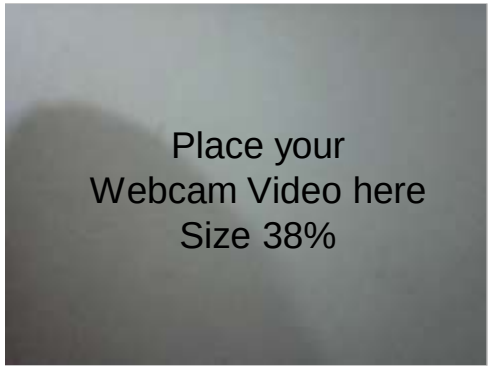
**Knowledge to be mined (or: Data mining functions)**

Characterization, discrimination, association, classification, clustering, trend/deviation, outlier analysis, etc.

Descriptive vs. predictive data mining

Multiple/integrated functions and mining at multiple levels

# <Content Slide>




## Techniques utilized

Data-intensive, data warehouse (OLAP), machine learning, statistics, pattern recognition, visualization, high-performance, etc.

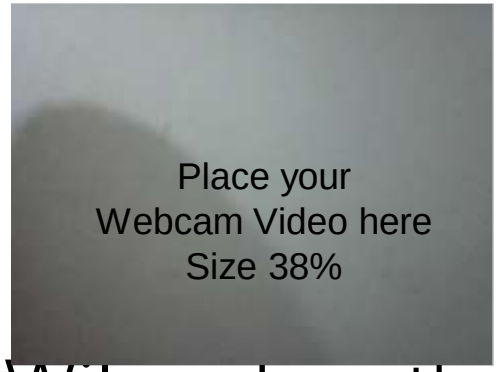
## Applications adapted

Retail, telecommunication, banking, fraud analysis, bio-data mining, stock market analysis, text mining, Web mining, etc.

# Summary

Data mining is popularly known as , knowledge discovery from data, or KDD  is the iterative sequence of the seven steps.

# Reference




1. William H Inmon “Building the Data Warehouse”, Wiley, Fourth Edition 2005.

2. K.P.Soman, Shyam Diwakar and V.Ajay, “Insight into data mining Theory and Practice”, Prentice Hall of India, 2006.

3. Jiawei Han, Micheline Kamber and Jian Pei, “Data mining concepts and Techniques”, Third Edition, Elsevier Publisher, 2006.