Introduction of Data mining

MCA V Sem

(Associate Prof.) Dr. Mahadev

Data Mining

- Definition
  - The process of extracting valid, previously unknown, comprehensible, and actionable information from large database and using it to make crucial business decisions
- Knowledge discovery
  - Association rules
  - Sequential patterns
  - Classification trees
- Goals
  - Prediction
  - Identification
  - Classification
  - Optimization

Data Mining Techniques

- Predictive Modeling
  - Supervised training with two phases
  - Training phase : building a model using large sample of historical data called the training set
  - Testing phase : trying the model on new data
- Database Segmentation
- Link Analysis
- Deviation Detection

What are Data Mining Tasks?

- Classification
- Regression
- Clustering
- Summarization
- Dependency modeling
- Change and Deviation Detection

What are Data Mining Discoveries?

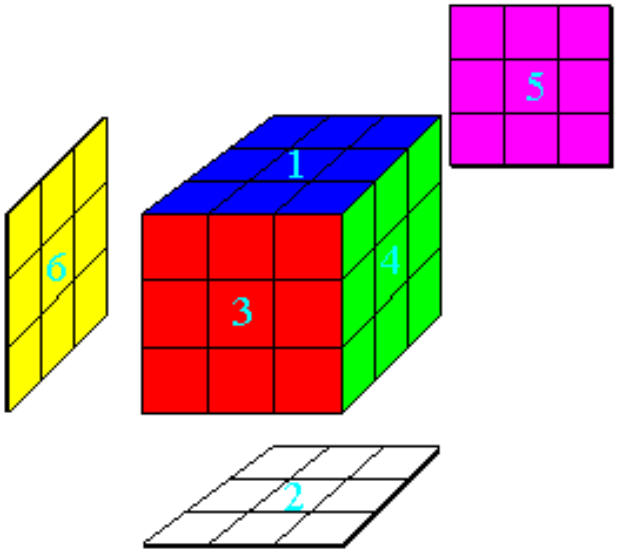
- New Purchase Trends
- Plan Investment Strategies
- Detect Unauthorized Expenditure
- Fraudulent Activities
- Crime Trends
- Smugglers-border crossing

Results of Data Mining Include:

- Forecasting what may happen in the future
- Classifying people or things into groups by recognizing patterns
- Clustering people or things into groups based on their attributes
- Associating what events are likely to occur together
- Sequencing what events are likely to lead to later events

Data Mining versus OLAP

- OLAP - On-line Analytical Processing
  - Provides you with a very good view of what is happening, but can not predict what will happen in the future or why it is happening



Data Warehouse for Decision Support & OLAP

- Putting Information technology to help the knowledge worker make faster and better decisions
  - Which of my customers are most likely to go to the competition?
  - What product promotions have the biggest impact on revenue?
  - How did the share price of software companies correlate with profits over last 10 years?

Decision Support

- Used to manage and control business
- Data is historical or point-in-time
- Optimized for inquiry rather than update
- Use of the system is loosely defined and can be ad-hoc
- Used by managers and end-users to understand the business and make judgements

Data Mining: Types of Data

- Relational data and transactional data
- Spatial and temporal data, spatio-temporal observations
- Time-series data
- Text
- Images, video
- Mixtures of data
- Sequence data

- Features from processing other data sources

Data Mining Techniques

- Supervised learning
  - Classification and regression
- Unsupervised learning
  - Clustering
- Dependency modeling
  - Associations, summarization, causality
- Outlier and deviation detection
- Trend analysis and change detection

Different Types of Classifiers

- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)
- Density estimation methods
- Nearest neighbor methods
- Logistic regression
- Neural networks
- Fuzzy set theory
- Decision Trees